Regionalisation of the parameters of the rainfall–runoff model PQRUT

Valeriya Filipova, Deborah Lawrence and Harald Klempe

ABSTRACT

This paper presents the regionalisation of the three parameter event-based PQRUT model, which is used for design flood analyses. The PQRUT model is used for the analysis of peak flows for which a sub-daily temporal resolution is required. The availability of high-resolution discharge data and disaggregated precipitation data have made it possible to re-evaluate the regional regression equations currently in use. We also assess whether the model parameters show spatial dependency. Event-based calibration was performed for the 45 highest flood events for each of 55 selected catchments across Norway, representing peak flows generated predominantly by rainfall. Due to the geographical heterogeneity of most areas in Norway, a statistically significant homogeneous region was only identified for catchments in southeastern Norway. Multiple linear regression and weighted regression were, therefore, used to develop a single set of equations, applicable to the entire country. The results for the weighted regression indicate a decrease in the median Kling–Gupta efficiency from 0.64 to 0.51 for calibration and regionalisation, respectively. The results also suggest that regression methods may perform better than methods based on spatial proximity in regions with varying topography when a parsimonious model is used.

Key words | PQRUT model, regionalisation, spatial autocorrelation, ungauged catchments

Valeriya Filipova (corresponding author)
Harald Klempe
University College of Southeast Norway,
    Department of Health and Environmental
    Studies (INHM)
N-3800 Bø,
Norway
E-mail: valeriya.filipova@hit.no

Deborah Lawrence
Norwegian Water Resources and Energy
    Directorate,
PO Box 5091 Maj.,
N-0301 Oslo,
Norway

INTRODUCTION

Significant advances have been made in the simulation methods used for estimation of design floods in recent years (Loukas & Kjeldsen 2013). This reflects both the development of newer simulation methods such as stochastic continuous and semi-continuous simulation (e.g., Blazkova & Beven 2004; Paquet *et al.* 2013) and the availability of more extensive data series for use in applying hydrological models to the problem of flood frequency estimation. However, the calibration of hydrological models is still required for such approaches, and the success of such methods is often dependent on the reliability of the hydrological model.

An additional challenge is that many design flood analyses are undertaken for ungauged catchments, such that regionalisation methods are needed to obtain values for the parameters of the rainfall–runoff models. These methods

transfer information from gauged catchments to ungauged or poorly gauged catchments (Blöschl & Sivapalan 1995). Razavi & Coulibaly (2013) distinguish two types of regionalisation: hydrological model independent vs dependent regionalisation. Hydrological model independent regionalisation is used to estimate values for streamflow directly from catchment characteristics, without the use of a hydrological model. Methods which relate hydrological model parameters to catchment characteristics and then estimate runoff with a hydrological model are classified as hydrological model dependent regionalisation. Regionalisation methods for ungauged catchments have been a widely researched topic for many years, and results summarising advances in prediction in ungauged basins are presented in Hrachowitz *et al.* (2013).

Commonly used methods for regionalisation include standard statistical methods, such as multiple linear regression, principal component analysis (PCA) and cluster analysis, as well as spatial statistical techniques such as kriging. Cluster analysis is a technique which is a well-suited and widely used method in regionalisation, as it has as its goal the identification of similarity. PCA is commonly performed prior to cluster analysis in order to reduce the dimensionality and to assign weights to the descriptors objectively (Burn & Boorman 1992). Cluster analysis can also be used to classify catchments prior to applying multiple regression (Burn & Boorman 1992).

When multiple regression is applied to the problem of regionalisation, a set of catchment descriptors is tested for their correlation with streamflow or with the calibrated parameters for a hydrological model estimating streamflow, and a corresponding regression model is built based on the catchment descriptors. If, however, a single set of regression equations is developed for a set of catchments with diverse geographical characteristics which are not fully described by the regression equations, the results can be rather poor (Merz & Blöschl 2004; McIntyre *et al.* 2005). On the other hand, Wagener *et al.* (2004) obtained good R^2 values for regression equations for the five parameters of a hydrological model as a function of catchment descriptors, based on a study of nine catchments in the UK. In cases where the hydrological model is described by many parameters (i.e., possibly overparameterised), however, some of them may not necessarily be significantly correlated with the catchment descriptors. For a relatively homogeneous region in Sweden, Seibert (1999) used the 11 catchments to regionalise the HBV-light model for monthly water balance and found that only six of the 13 model parameters were significantly correlated with catchment descriptors. An additional disadvantage of regression models is the assumption underlying regression analysis for estimating model parameters that they are independent, which is often not the case for heavily parameterised hydrological models (McIntyre *et al.* 2005).

Methods which take account of spatial information include ordinary kriging, top kriging and nearest neighbour approach. In addition to these methods, the parameters of the catchments immediately downstream and upstream can be transferred to the ungauged catchment (Merz &

Blöschl 2004). Ordinary or top kriging is a method in which the model parameters are interpolated. Then, the model parameters at the catchment centroids are read directly from the map (Viviroli *et al.* 2009). The advantage of top kriging compared to ordinary kriging is that the correlation in the runoff between upstream and downstream catchments is considered and the kriging weights take into account the catchment size (Vormoor *et al.* 2011).

Uncertainty needs to be considered in regionalisation studies because the errors in input data, model structure and calibration are in most cases significant. A common method to propagate these errors into the resulting relation between model parameters and catchment descriptors is to use Monte Carlo simulations. For example, McIntyre *et al.* (2005) uses ensemble modelling, based on the generalised likelihood uncertainty estimation, to regionalise the parameters of the probability distribution model, using 127 catchments in the UK. In a study by Wagener *et al.* (2004), a weighted regression was used to regionalise the five parameter pd4–2pll model, representing the probability distributions of stores and routing components. The proposed framework involved using a weighted regression, based on the 'identifiability' of the model parameters. The identifiability was estimated by using the gradient of the cumulative distribution of the parameters for different segments of the parameter space. Catchments with high identifiability received higher weights in the multiple regression. The results of the study showed an improvement in the coefficient of determination for some of the parameters when weighted least squares regression is used compared to using ordinary least squares (OLS) regression. However, similar Nash–Sutcliffe efficiency (NSE) values of 0.78 were achieved using the two regression methods.

As can be seen from this summary of relevant studies, different levels of effectiveness are reported for the regionalisation methods applied. This most likely reflects physiographic differences between the locations used as study sites, as well as differences in data availability and quality. As it can be very difficult to determine which method will perform best for a given locality, there is a strong argument for testing alternative methods for regionalisation before selecting a preferred method.

In this study, model dependent regionalisation will be used to generalise the parameters for a simple, event-based

model, PQRUT, used for estimation of design floods in Norway. PQRUT is an event-based model with three parameters, which are, in principle, linked to the shape of the hydrograph for high flow events. These parameters can be calibrated where data are available, although PQRUT is primarily used for ungauged catchments, such that other methods are required for setting the three parameters. The parameter estimates currently used for ungauged catchments are based on a regionalisation using regression equations developed over 30 years ago (Andersen *et al.* 1983). For that regionalisation, 38 catchments were considered, representing sites for which sub-daily rainfall data were available from meteorological stations equipped with pluviometers. All of these sites were located in the southern half of Norway. Hourly values of rainfall were extracted from pluviographs from the nearest meteorological station for comparison with high-resolution discharge data for a given catchment. As the PQRUT model has no parameters to account for storage prior to or changes in storage during an event, full saturation was assumed at the start of each event used for calibration. In order for this assumption to be satisfied, only discharge events which were preceded by a cumulative precipitation of at least 50–100 mm in the two preceding weeks were used in the calibration. Following the calibration of individual sites, stepwise multiple linear regression was used to select significant catchment descriptors and derive regional equations. In addition, recommendations for adjustments of the K_1 parameter values, depending on the effective lake percentage and the fraction of marsh in the catchment, were given based on the results of the analysis.

During the intervening time period, the availability of both discharge and precipitation data with high temporal resolution has increased significantly, providing much needed input data to support the 1-hour time step usually used in model applications. In this work, we therefore re-evaluate the empirical equations using newer precipitation and discharge data series and a wider range of catchments consider methods for incorporating uncertainty into the parameter estimates given by the regression equations, and also test for the presence of spatial autocorrelation in the parameters. This latter factor is of interest as it could support the use of kriging or parameterisation by sub-regions. As there is a continued need for extreme flow estimates at ungauged sites, a re-evaluation of the original equations based on newer data and methods is long overdue.

METHODS

Study area and data set

Study catchments were primarily selected using the following two criteria: the streamflow in the catchments must be unregulated, the catchments must have at least ten years of hourly discharge data. In addition, it was necessary that a calibrated HBV model already existed for the catchment (Lawrence *et al.* 2009), as this assisted in selecting rainfall-dominated discharge events for calibration, as further described in the following section. The catchment size was restricted to small- and medium-sized catchments, as practical applications tend to indicate that the PQRUT model overestimates peak values in larger catchments. This is most likely due to the simple model formulation which does not allow for internal storage during a runoff event. Although it is difficult to establish an upper boundary, as this will vary significantly with catchment characteristics and flood generating processes, the largest selected catchment in this study is 1,648 km^2 while the median size is 170.5 km^2 (Table 1). Figure 1 illustrates the 55 catchments used for the calibration, and although they are generally fairly well distributed across Norway, the northern region has fewer catchments such that the catchment density is particularly limited in that region. The selected catchments are all members of the Norwegian Bench Mark data set (Fleig *et al.* 2013), ensuring that the data series are of sufficient quality for modelling, at least, daily discharge, including flood peaks.

Catchment descriptors related to geomorphologic: area (*A*), mean altitude (*Hm50*), length (*L*), effective lake per cent (*Lk*), drainage density index (*DrInd*) and catchment steepness (*Hl*); land cover: fraction of forest (*For*), fraction of marsh (*M*), fraction of sparse vegetation over tree line (*B*); climatic and hydrologic properties: mean temperature in the summer (*Ts*), mean temperature in the winter (*Tw*), mean annual temperature (*Temp*), mean annual runoff (*Q*) and mean annual precipitation (*P*), were selected for analysis (Table 1). The catchment steepness is defined as

Table 1 | Minimum, maximum, median, mean, standard deviation and coefficient of variation for the selected catchment descriptors

	Min	Max	Median	Mean	Standard deviation	Coefficient of variation
Area (A), km^2	60	1,648	170.5	290.9	315	1.1
Mean annual streamflow (Q), mm/yr	444.8	4,817.9	1,315.8	1,567.1	1,004.9	0.6
Mean annual precipitation (P), mm	510	4,907	1,667	2,057.5	1,127	0.5
Median elevation ($Hm50$), m	192	1,349	760	746.2	340.8	0.5
Drainage density index ($DrInd$), 1/km	0.8	2.7	1.8	1.8	0.5	0.3
Length (L), km	2.4	74.8	21.5	25.1	15.3	0.6
Fraction of forest (For)	0	0.9	0.3	0.4	0.2	0.7
Fraction of marsh (M)	0	0.3	0	0.1	0.1	1
Fraction of sparse vegetation over tree line (B)	0	0.9	0.4	0.4	0.3	0.7
Effective lake percentage (Lk)	0	0.1	0	0	0	1.2
Catchment steepness (Hl), m/km	2.8	79.1	12.2	17.9	18	1
Mean temperature in winter (Tw), °C	−7.6	2	−2.8	−2.9	2.6	−0.9
Mean temperature in summer (Ts), °C	5.3	12.9	8.3	8.7	1.9	0.2
Mean temperature ($Temp$), °C	−2.3	6.2	1.6	1.7	2.1	1.3
K_1	0.0088	0.3039	0.0723	0.0871	0.0614	0.7042
K_2	0.0023	0.137	0.0134	0.0181	0.0195	1.0786
T	14.4	79.4	42.3	43.3	12.5	0.3

H_{75}-H_{25}/L, where H_{75} and H_{25} are the elevation at the 75 and 25% levels of the hypsometric curve. There is much variation in the catchment descriptors for the 55 catchments (Figure 2). The area of the catchments varies from 6 to 1,648 km^2, the mean annual runoff from 494 to 4,885 mm/yr, and the mean elevation is between 192 and 1,538 m. The catchments were delineated and the catchments descriptors, excepting the mean annual runoff, precipitation and temperature, were obtained from tools available in the NVE atlas (www.nve.no). The mean annual streamflow data were calculated by using the data sets for daily discharge for the entire record length, which also varied somewhat between catchments. The temperature and precipitation were calculated for the normal period for 1961 to 1990 from the 1 by 1 km interpolated gridded data (www.seNorge.no), available for precipitation and temperature in Norway (Mohr 2008).

PQRUT

The PQRUT model is an event-based model that follows a simple reservoir concept (Figure 3) related to the shape of the hydrograph. The model is primarily applied to ungauged or regulated catchments for which it is not possible to calibrate a more complete hydrological model, such as HBV. The parameters K_1 and K_2 represent the fast and slow response of the hydrograph, and the parameter T is the threshold value above which the fast response parameter K_1 is activated. The following formulas (Andersen *et al.* 1983) are currently used to estimate the three parameters if the model cannot be calibrated:

$$K_1 = 0.0135 + 0.00268\, Hl - 0.01665\, lnA_{SE} \qquad (1)$$

$$K_2 = 0.009 + 0.21\, K_1 - 0.00021\, Hl \qquad (2)$$

$$T = -9.0 + 4.4\, K_1^{-0.6} + 0.28\, QN \qquad (3)$$

where Hl is the catchment steepness, A_{SE} is the effective lake per cent, and QN is the 'normal' specific runoff, which in practice is estimated for the period 1961–1990 from seNorge runoff data for the entire country.

The model is used, among other things, for the estimation of design floods and safety check floods for dam

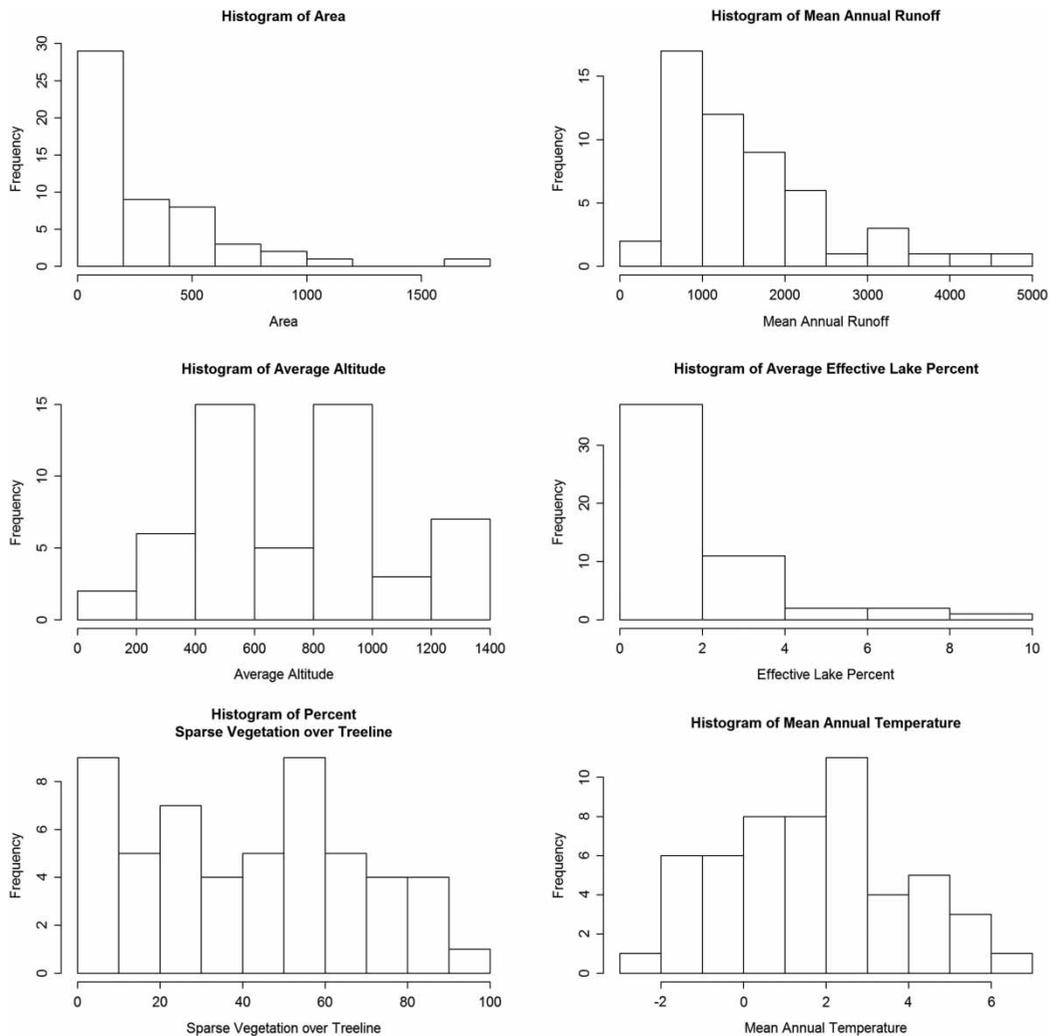**Figure 1** | Location of the 55 study catchments in Norway.

Figure 2 │ Distribution of catchment descriptors for the catchments shown in Figure 1.

design in Norway. A hypothetical precipitation design sequence of a given return period is routed through the PQRUT model, usually under the assumption of full catchment saturation. Depending on the season most vulnerable to extreme flood events, an additional volume can be added to the input sequence to account for a snowmelt contribution (see Lawrence et al. (2014) for further details and a comparison with other methods). The PQRUT model parameters are usually estimated by the equations given above, but infrequently, calibration of the parameters has been previously achieved for catchments during practical applications using observed precipitation and discharge series. The resulting design discharge sequence is given by the PQRUT model output, and in dam safety applications,

this sequence is then further routed through a reservoir to given design water levels.

Event selection and calibration

In order to calibrate an event-based model, individual discharge events must be selected from the observed time series. For each catchment, the 45 highest discharge events were selected from the 1-hour time series using several criteria. First, the duration of the high flow event must be identified, and this was done based on an estimate for the baseflow. The start and the end of the event was chosen to be higher than 1.05 of the baseflow, as determined by using the Lyne and Hollick filter (Lyne & Hollick 1979),
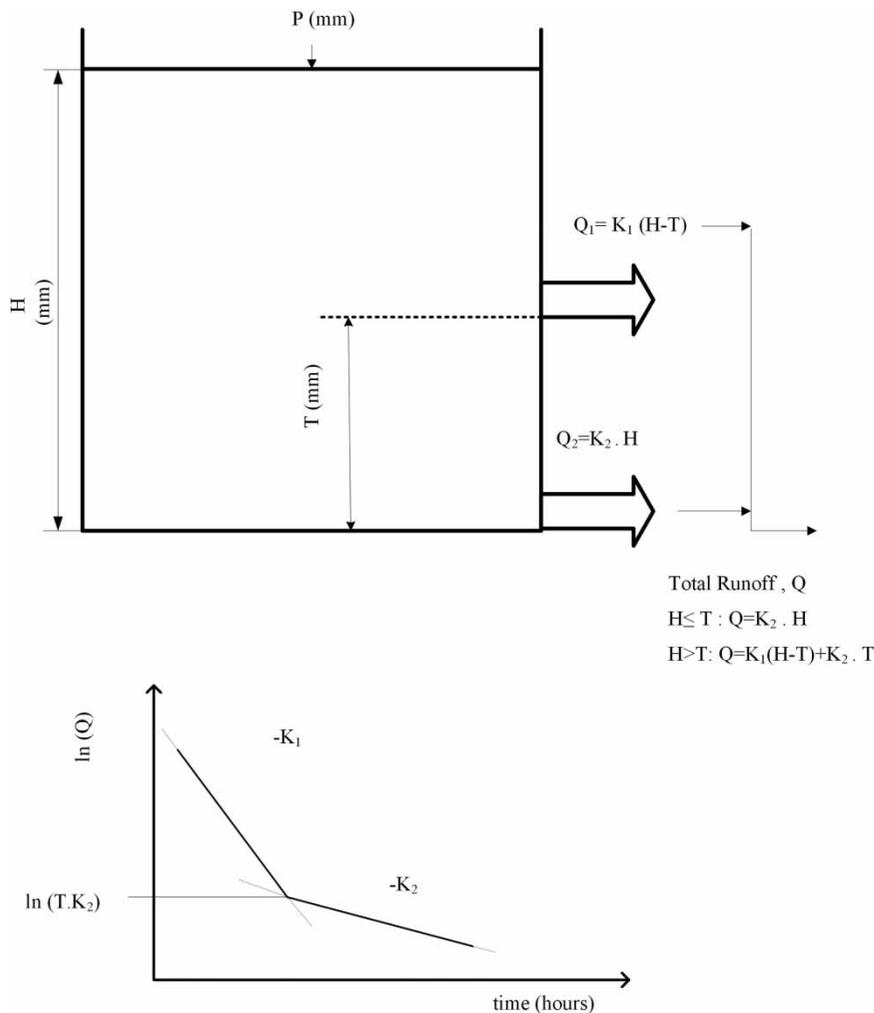
Figure 3 │ Model structure of PQRUT.

implemented using an R package (Ladson et al. 2013). Recursive digital filters are commonly used as a baseflow separation method (e.g., Serinaldi & Grimaldi 2011; Gaal et al. 2015). The Lyne and Hollick filter is a digital filter, originally used in signal processing. As such, this technique does not consider the underlying physical processes when separating direct runoff and baseflow. However, the Lyne and Hollick filter can be considered for this application because only the start and end of the event need to be determined. The filter is based on a recursive method such that the equation for the direct runoff (Qd) at time i can be written as: $Qd(i) = \alpha Qd(i-1) + 1 + \alpha/2[Q(i) - Q(i-1)]$, where α controls the decrease in the percentage baseflow for each pass through the data. The minimum number of passes is

three (forward, backward and forward), but can be increased to further refine the baseflow adjustment. In this application, the standard value of 0.925 of the parameter was used and the procedure was set to execute 11 passes through the time series for each catchment. Following this filtering, individual events could be extracted for potential use in the PQRUT calibration.

It was also important to select events for the calibration which are primarily driven by rainfall, rather than combined rainfall and snowmelt. Although the latter are important in design flood applications, the lack of a suitable model for hourly snowmelt in the wide range of catchments considered here makes it difficult to correctly specify well-defined hourly input for the calibration. The results from a

calibrated HBV model for each catchment were, therefore, used to obtain values for the snowmelt for the duration of the flood event. Flood events that had less than a 20% snowmelt contribution during the entire flood event were selected for calibration.

In addition to the criteria for event duration related to the baseflow filtering, an independence criteria of 350 hours was imposed on the events selected for each catchment. The selected events were then also screened visually, to ensure the quality of the sub-daily time series. Hydrographs suggesting poor data quality, for example, having large steps in the hourly discharge, were discarded. In some catchments, events selected using the baseflow criterion had long durations (>500 hours), particularly if they occurred during low baseflow periods. In these cases, the hydrographs were truncated manually based on the visual inspection. This final process of manual data screening was very time-consuming, but was found to be necessary for obtaining good calibration results from the hourly discharge data.

The time step used to run the model was set to 1 hour. Precipitation data with a 3-hour temporal resolution, representing a temporal disaggregation of daily gridded data using the HIRLAM model (Vormoor & Skaugen 2013), were used to drive the model. A uniform distribution was used to further disaggregate the rainfall data to a 1-hour time step. The events were calibrated by using the dynamically dimensioned search optimisation (Tolson & Shoemaker 2007) available in ppso package in R, and the Kling–Gupta efficiency (KGE) criterion (Gupta *et al.* 2009) was used as the objective function.

In addition to the three PQRUT parameters illustrated in Figure 3, an additional parameter, lp, was introduced to account for initial losses to the soil zone, depression storage and evapotranspiration at the onset of the event. Although model applications for design floods generally assume full saturation, this is not necessarily the case for the events selected from the time series, particularly after the bulk of events involving snowmelt have been excluded. In addition, the exclusion of all hydrographs for which initial conditions are less than fully saturated, would lead to too few representative events for model calibration. Therefore, this additional parameter was introduced so that a wider range of events could be included in the calibration relative to

those which would be identified if a constraint of full saturation (i.e., zero storage capacity) were imposed, such as was done in the Andersen *et al.* (1983) calibration. This parameter functions as an initial loss to the system, such that the input precipitation to the reservoir model (Figure 3) is 0 until the value of lp is exceeded by the cumulative input rainfall. The value of the initial loss, lp, was calibrated for each catchment and for each flood event. The ranges of the parameter lp established during the event calibrations are given in Table 2. As the parameter lp was introduced as a tool to facilitate calibration using a wider range of peak flow events, it was not regionalised. Furthermore, as the return periods for events estimated in a design flood analysis tend to be relatively long in comparison with those of the events considered in the calibration, an assumption of full saturation (i.e., lp = 0) is often justified. In further work, stochastic modelling methods, such as semi-continuous simulation (Paquet *et al.* 2013) could be used to determine suitable probability distributions for such initial losses for long return periods.

The starting point used for the optimisation for the calibration of the model parameters was the parameters' values estimated from the previously developed regression equations (i.e., Equations (1)–(3)). If the calibration for an event achieved a KGE value of less than 0.7, it was calibrated again using a random starting point. Only calibrated parameter values corresponding to KGE values greater than 0.7 for a given catchment were retained for further use in the regionalisation analyses. Outlying parameter values were detected using an algorithm based on principal components, available from the mvoutlier package in R (Filzmoser *et al.* 2008). The outliers were excluded and the remaining parameter values for the calibrated events, if there were more than five, were averaged. The catchments for which calibration was not successful were removed

Table 2 | Parameter ranges used for calibration

Parameter	Range	
K_1	0.001	0.35
K_2	0.001	0.35
T	1	100
lp	0.01	50

from the analysis. Due to poor model performance (i.e., there were less than five events with KGE >0.7), six of the catchments were excluded from the analysis. These catchments had either a high percentage of glacier cover (>20%) or a snowmelt dominated flood regime.

Multiple regression

The PQRUT parameters were regionalised using stepwise multiple linear regression. The assumptions underlying linear regression include: a linear relationship between the dependent and independent variables, a constant error variance (homoscedasticity), and errors which are independent and are normally distributed (Kottegoda & Rosso 1997). The multicolinearity of the catchment descriptors increases the uncertainty in the regression model because different values of the observation can produce the same fit and the regression model can become unstable. The functions vif and outlierTest from the package car in R were used to perform diagnostic tests for the multiple regression (Fox & Weisberg 2011). Variance inflation factors are used to determine if there is colinearity in the variables in the linear regression. The forward and backward AIC was used to select independent variables. The Bonferroni test was used to detect the outliers of the regression and the Shapiro–Wilk test was used to determine if the residuals of the regression are normally distributed. The outliers and leverage points were also detected visually using the Cook's D plot. Scatter plots of the parameters and the catchment descriptors, selected in the regression equations, were examined for non-linear correlation. In cases of non-linear correlation, several transformations of the variables were used and the value that maximised the linear correlation was selected. The adjusted R^2 was calculated using the Wherry formula in R, which takes into account the number of independent variables used in the regression equation.

In order to account for the uncertainty in the calibration, the inverse of the standard error for the mean of the parameters was used as weights in a weighted regression. Using this approach, catchments that have a high standard deviation and a low number of successfully calibrated events (i.e., KGE >0.7) have lower weight in the regression analysis. The weights were standardised by

dividing by the mean. This approach is similar to the method used by Wagener *et al.* (2004), regionalisation using weighted regression taking into account the identifiability of the model parameters.

To further analyse the uncertainty in the calibration, another set of equations was derived for the coefficient of variation of the parameters. These equations correlated the coefficient of variation of the calibrated parameter values with the catchment descriptors. The equations can provide the practitioner with information concerning the type of catchments for which the calibration of the parameters resulted in a larger spread of values for the individual events, suggesting a higher degree of uncertainty in practical application.

The model was validated by leave-one-out cross validation (LOOCV). Although this method of validation can result in inconsistencies, the LOOCV was chosen so that maximum information can be retained in the analysis. If the set of catchments was divided into, e.g., a two-thirds training set and a one-third validation set, then only 33 catchments could be used instead of the full set of 49 for deriving the regression equations. The LOOCV method, in contrast, allows one to use all of the catchments. After excluding outliers, one catchment was retained (each time) for validation, and the regression coefficients of the equations were estimated by using the remaining catchments. The parameters were then estimated for this catchment, using the coefficients of the resulting equations. The PQRUT model was then applied using these parameters for the validation catchment for all of the events with successful calibrations. The procedure was repeated for all catchments. The median of the KGE values from the events was used as a criterion for the comparison of the performance of three different models: (1) the existing equations currently in use based on Andersen *et al.* (1983), (2) OLS, and (3) weighted regression.

Spatial patterns and spatial autocorrelation of the parameters

The Moran's I test was performed on the calibrated parameters and on the residuals from the linear regression to test for spatial autocorrelation. High positive values of I indicate spatial clusters, and negative values indicate that the

features are scattered (Mitchell 2005). The Moran's I test was performed using the spatial statistics toolbox in ArcGIS. In addition, partitioning around the medoids (PAM) cluster analysis was used to determine whether geographically homogeneous regions in parameter values and residuals can be identified. The cluster analysis depends on the dissimilarities of the objects. For the PAM cluster analysis, a set of representative objects or initial cluster medoids are selected and then the remaining objects (catchments) are assigned to the nearest medoid based on Euclidian distance, calculated from the selected attributes (Kaufman & Rousseeuw 1990). The medoids of the clusters are then recalculated and the objects are assigned again to the nearest medoid (Kaufman & Rousseeuw 1990). The algorithm continues until none of the objects change cluster membership (Kaufman & Rousseeuw 1990).

RESULTS

Correlation between parameters, catchment descriptors and coefficient of variation

The results from calibrating the PQRUT model for individual catchments showed high coefficients of variation and skewed distributions, even after the outliers have been removed prior to calculating the mean values (Figure 4). Due to poor model performance, six of the catchments were excluded. These catchments had a high glacier percentage (>20%) or a snowmelt-dominated flood regime such
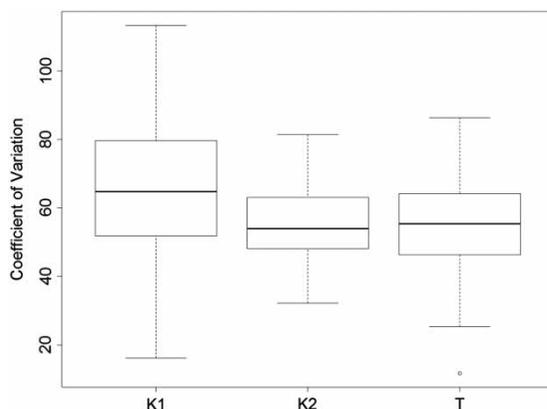

Figure 4 | Coefficient of variation, expressed as percentage, for the calibrated parameters.

that the availability of too few events undermined the model calibration. After exclusion of these catchments, the first step for developing the regression equations was to investigate the correlation between the parameters and the catchment descriptors (Table 1). The Spearman rank correlation, unlike the Pearson correlation, is non-parametric and does not require a linear relationship between the variables but does assume a monotonic relationship (Wagener *et al.* 2004). For this reason, the Spearman rank correlation is preferred (e.g., Seibert 1999; Wagener *et al.* 2004; Skaugen *et al.* 2014). The Spearman rank correlations are shown in Table 3.

The parameter K_1 (representing the fast runoff rate, Figure 3) is significantly correlated with the effective lake percentage (*Lk*), having a Spearman correlation of 0.69, and the mean annual precipitation (*P*). The parameter K_2 (slow runoff rate) is significantly correlated with all of the catchment descriptors (*A, Q, P, Temp, Tw, Hl, Lk* and *DrInd*), except for those related to land cover (*M, B, For*), the summer temperature *Ts* and the median elevation *Hm50*. The correlation coefficients for the significantly correlated descriptors are between 0.32 and 0.44. The high correlation with temperature (*Temp* and *Tw*) occurs because temperature is also significantly correlated with *Hl*, the relief parameter, and *Q*, the average annual discharge. This is a spurious correlation reflecting a co-occurrence of warmer temperatures, high annual rainfall and pronounced topography in western Norway. The parameter T (representing the threshold between the fast and slow runoff rates) is significantly correlated with: mean annual streamflow (*Q*), precipitation (*P*) and the effective lake percentage (*Lk*). The parameters of the model are also strongly correlated. The Spearman correlation between K_1 and K_2 is 0.54 and between K_1 and T is 0.49.

The results for the Pearson correlation between the coefficient of variation and the catchment descriptors are given in Table 4. The coefficient of variation for K_1 is significantly ($p < 0.05$) correlated with most of the catchment descriptors, although the coefficient of variation for K_2 is significantly correlated only with area (*A*), and T only with the catchment length (*L*). The results for the correlation of the coefficient of variation of the parameters and the catchment descriptors indicate that the coefficient of variation for K_2 and T is lower for catchments with larger area.

Table 3 | Spearman correlation between the model parameters and the catchment descriptors

	A	Q	P	Hm50	Drind	L	For	M	B	Lk	Hl	Tw	Ts	Temp	K_1	K_2	T
K_1	0.06	−0.27	−0.32	0.06	−0.1	0.02	0.04	0.15	0	−0.69	−0.03	−0.22	−0.03	−0.15	1	0.54	−0.49
K_2	−0.37	0.32	0.33	0	0.33	−0.42	−0.03	−0.1	0.08	−0.33	0.44	0.35	0.14	0.32	0.54	1	−0.39
T	−0.04	0.39	0.32	0	0.11	0.04	−0.13	−0.16	0.08	0.5	0.04	0.17	−0.02	0.11	−0.49	−0.39	1

Table 4 | Pearson correlation between the coefficient of variation of the parameters and the catchment descriptors

	A	Q	P	Hm50	Drind	L	For	M	B	Lk	Hl	Tw	Ts	Temp	K_1cv	K_2cv	Tcv
K_1cv	0.41	−0.52	−0.51	0.18	−0.45	0.43	0.11	0.22	−0.16	−0.31	−0.31	−0.54	−0.14	−0.45	1	−0.07	−0.2
K_2cv	−0.29	0.01	0.08	−0.12	0.04	−0.23	0.28	−0.01	−0.19	−0.13	0.19	0.14	0.22	0.18	−0.07	1	0.36
Tcv	−0.23	0.02	0.02	−0.13	0.13	−0.32	0.19	0.15	−0.17	0.02	0.11	0.09	0.19	0.14	−0.2	0.36	1

Resulting regression equations

Regression equations are developed for estimating the three model parameters, using OLS regression and weighted regression. Another set of OLS regression equations is developed for the standard deviation of the parameters. The results from the correlation analysis were used for initial selection of catchment descriptors for the stepwise multiple regression, also referred to here as OLS regression. Many of the catchment descriptors are correlated, and for this reason, the catchment descriptors that had the highest correlation with the parameters from each group were selected. The resulting equations are:

$$K_1 = 0.031 - 0.0343 \text{ For} - 1.35.10^{-5}P - 0.0102 \text{ lnLk} + 1.83\ K_2 \tag{4}$$

$$K_2 = -0.031 + 0.000521 \text{ Hl} - 0.004184\ \text{lnLk} + 0.01026 \text{ DrInd} \tag{5}$$

$$T = 7.31 + 0.00636\ P + 1.591\ K_2^{-0.6} - 0.5343 \text{ lnLk} - 0.09283 \text{ L} \tag{6}$$

Two catchments were removed as outliers for the equation for K_1, one of these having a very low percentage *For* (forest cover), such that the dominant land cover is sparse vegetation over tree line (i.e., *B* is 0.8). The other catchment is the largest in the data set with an area of 1,648 km^2, with low *P* of 626 mm/yr. The outliers for the equations for K_2 were two catchments that have very high *Hl* values (over 60). The catchment that was an outlier for the equation for T is fairly elongated (having a length of 75 km for a total catchment area of 854 km^2) and has a large percentage of marsh at 14%. These two factors induce a particularly high T value.

For the weighted least squares regression, the same dependent variables were used as had been used for the OLS regression. The resulting equations are:

$$K_1 = 0.00985 - 0.0225 \text{ For} - 1.04.10^{-5}P - 0.00996 \text{ lnLk} + 1.91\ K_2 \tag{7}$$

$$K_2 = -0.023169 + 0.000497\ Hl - 0.003491 \text{ lnLk} + 0.007229 \text{ DrInd} \tag{8}$$

$$T = 9.66 + 0.00728\ P + 1.81716\ K_2^{-0.6} +\ 1.90470 \text{ lnLk} +\ 0.00721\ L \tag{9}$$

The same catchments identified and removed as outliers for the OLS regression were also identified and removed as outliers for equations (Equations (7) and (8)) for K_1 and K_2. For Equation (9) for T, three catchments were removed as outliers.

The resulting equations for the coefficient of variation for the parameters are:

$$K_1cv = 88.24 - 0.016\ Q - 3.97 \text{ DrInd} +\ 0.018 \text{ Area} \tag{10}$$

$$K_2cv = 67.86 - 3.74 \text{ ln Area} +\ 14.99 \text{ For} \tag{11}$$

$$Tcv = 78.75 - 5.46 \text{ ln Area} + 48.80\ M \tag{12}$$

Two outliers were detected for Equation (10) and removed, and no outliers detected for Equations (11) and (12). The coefficients of determination for equations (Equations (4)–(12)) are given in Table 5.

Spatial pattern

The results from Moran's I test, performed for the parameters of the PQRUT model and for the residuals of the OLS regression, using distance threshold of approximately 250 km^2 indicate that the spatial autocorrelation for cluster pattern is not significant ($p < 0.05$) (Table 6).

Table 5 | R^2, adjusted R^2 and residual mean square error (RMSE) for the regression equations

Equations for:	R^2	Adjusted R^2	RMSE	Outliers
K_1	0.785	0.764	0.0284	2
K_2	0.701	0.679	0.00669	2
T	0.494	0.443	9.53	1
K_1 weighted regression	0.813	0.795	0.0195	1
K_2 weighted regression	0.59	0.561	0.00526	2
T weighted regression	0.744	0.717	6.88	3
K_1cv	0.539	0.507	16.2	2
K_2 cv	0.251	0.217	9.94	
Tcv	0.203	0.167	13.7	

Table 6 │ Moran's I values for the mean parameter values and the residuals of the regression equations for each parameter

	K_1	K_1 residuals	K_2	K_2 residuals	T	T residuals
Moran's I	−0.077	−0.156	−0.019	0.067	−0.023	−0.085
z-score	−0.958	−1.744	0.003	1.182	−0.059	−0.818
p	0.338	0.081	0.998	0.237	0.953	0.413

In addition, PAM cluster analysis was performed on the four catchment descriptors (*DrInd*, *Hl*, *Lk* and *Q*) that both show highest correlation to the parameters and also contain information about the regional pattern. The resulting classification performed for the 49 catchments did not show geographically homogeneous regions, except for the catchments which belong to cluster 2, in southeast Norway (Figure 5).

The highest number of catchments is classified to cluster 1 and includes catchments that are located mostly in mid-Norway but also some catchments located in southern Norway or western Norway. The resulting clusters 3 and 4 contain small catchments, located mostly in south western Norway. The number of catchments in clusters 3 and 4 is under 10, which means that robust regression equations could not be developed.

Performance of the regionalisation methods

In order to test the performance of the new equations, a comparison was made between estimates based on the original calibration from 1983 (Equations (1)–(3)), the average parameter values from the calibration for each catchment, and for the LOOCV based on the calibrated values. The results for the median KGE of the simulated flood events are shown in Figure 6. The parameters were then estimated by using the regional equations (Equations (7)–(9)) directly to obtain estimates for the selected catchments, including the outliers and plotted in Figure 7. For this simulation, where fully saturated conditions were assumed for all events, the highest results are given by the mean values from the local calibration of the different flood events: median KGE is 0.64, for weighted and OLS regression, the median KGE is 0.51 and 0.49 and the lowest values are from Equations (1)–(3): 0.46. The results for the simulations are given in Table 7.

DISCUSSION

Coefficient of determination for the equations

There is significant correlation between the three model parameters and some of the catchment descriptors. Regression equations that provide robust estimates could therefore be developed for the parameters of PQRUT. Similar results were found by Lee *et al.* (2006) for a five-parameter model applied to non-urban catchments with a low permeability in the UK. In that study, the reported loss in the NSE due to using parameters estimated from the regression equations is only 0.06 when compared with the local calibration. The loss in the NSE was higher for catchments with a higher baseflow index (BFIHOST >0.5). In addition, a study by Skaugen *et al.* (2014) shows that all the parameters of the seven calibrated parameters of the model distance distribution dynamics used in Norway are significantly correlated with the catchment descriptors. The median NSE for the 17 validation catchments for the data series for the validation was 0.66 compared to 0.78 for the validation of the calibration. For 10 of the 17 catchments, the deviations from NSE were less than 0.1.

Spatial proximity

No significant spatial autocorrelation in the calibrated model parameters was found in this study. This result is in contradiction with a number of studies, for example, Merz & Blöschl (2004), which shows that geostatistical methods perform better than regression methods. A reason for the lack of a statistically significant spatial autocorrelation is that the parameters K_1, K_2 and T are strongly correlated with geomorphologic catchment descriptors such as *Hl*, *L*, *Lk* and *DrInd*. However, the results of the Moran's I test might have been affected by the use of a relatively small
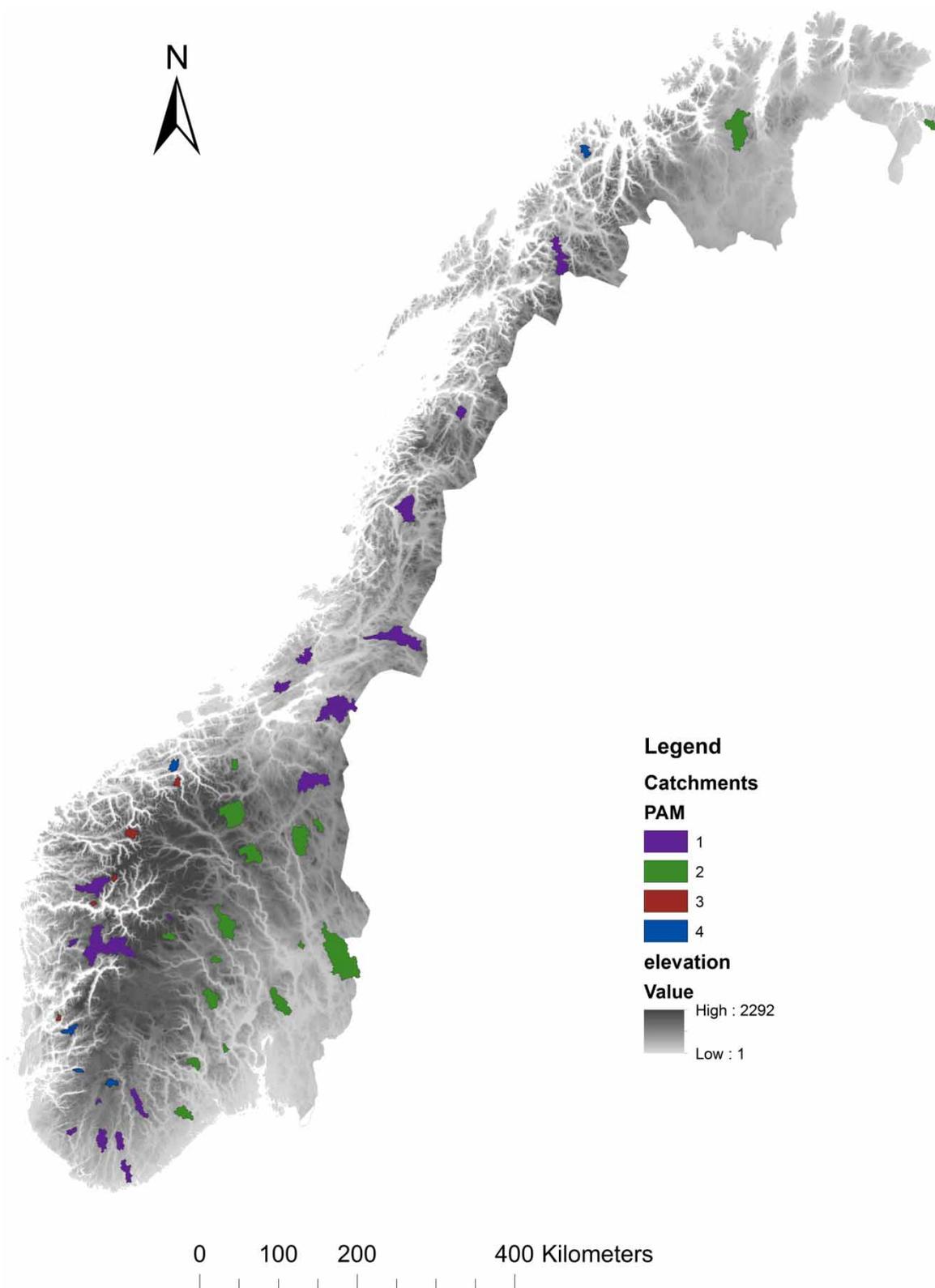
Figure 5 │ PAM cluster analysis for the selected catchments, based on *Q*, *Hl*, *Lk* and *DrInd* for the catchments.
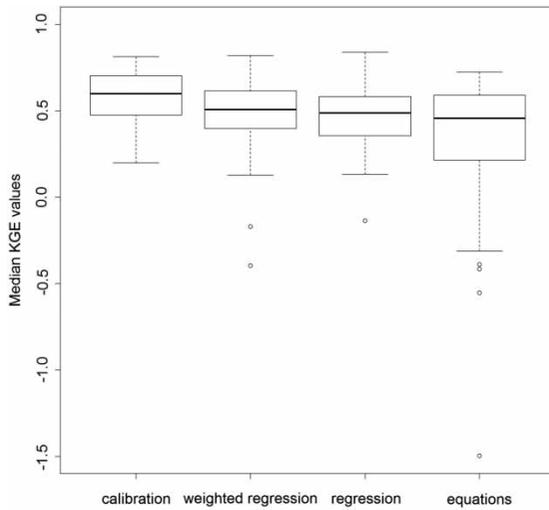
Figure 6 | Median KGE values for the streamflow simulations, using values, estimated by: the mean parameters from calibration at each catchment, Equations (1)–(3), regression and weighted regression.

data set of catchments that included few neighbouring and nested catchments. The lack of spatial autocorrelation precludes the use of kriging or other methods that take into account spatial proximity. The low spatial autocorrelation of the residuals shows that the residuals are spatially independent and it is acceptable to use multiple regression. Another reason for the low spatial autocorrelation is that the parameters K_1, K_2 and T are strongly correlated with geomorphologic catchment descriptors such as *Hl*, *L*, *Lk* and *DrInd*.

Catchments with low performance of calibration and regionalisation methods

The catchments for which lower calibration results are obtained (here defined as the median KGE < 0.5), have Q values (median Q = 670 mm/yr) that are lower than the median Q (1,316 mm/yr) for all catchments. If the performance of the calibration and regression methods is analysed for the 'wetter catchments', defined here as catchments that have a higher Q than the median Q for all catchments, the median KGE is 0.61 for the weighted regression, 0.56 for the OLS regression and 0.7 for the calibration. These values are higher than if the results from all the catchments are considered. A study by Merz & Blöschl (2005) that compares the use of spatial proximity methods with regression for

regionalisation for flood frequency analysis also shows that regionalisation methods perform better in wetter catchments. They define wet catchments as catchments with mean annual flood (MAF) higher than the median MAF for all catchments. The catchments where the performance of the calibration and regionalisation is lower are located in areas where Q and P are lower, especially in southeast Norway (Figure 7). A comparison between the design flood values that resulted from applying flood frequency analysis and from PQRUT also shows that the values estimated from the PQRUT are possibly overestimated in southeast Norway (Wilson *et al.* 2011). A possible solution is to develop a separate set of equations in this region. The lower values of P also make the assumption of fully saturated initial conditions invalid for many of the selected flood events, particularly during warmer periods.

Parameter uncertainty

There is a high degree of correlation between the parameters and this affects the equifinality of the model, although there are only three model parameters. High uncertainty has also been reported for the simple two-parameter unit hydrograph model (Yeh *et al.* 1997). In addition, there is uncertainty in defining the initial conditions which is a problem common for event-based models. The K_1 parameter is activated only if the storage is above the threshold, and below this value the parameter search is insensitive to the direction of K_1 because the parameter is not used in the simulation. Another reason for the uncertainty is that because PQRUT is an event-based model, the optimisation is performed on few data points (most of the selected flood hydrographs have around 200 hourly time steps). Due to this, the calibration results are unstable.

The correlation between the coefficient of variation and the catchment descriptors provides some insight into the reasons for the uncertainty in the estimated mean parameter values. However, it needs to be realised that because the parameter values are correlated, errors in one of the parameters is compensated by the other parameter values. The coefficient of variation for K_2 and T show positive correlation with *Hl* and *For* and negative correlation to *A*. These results indicate that in small catchments, there is higher uncertainty in simulating the
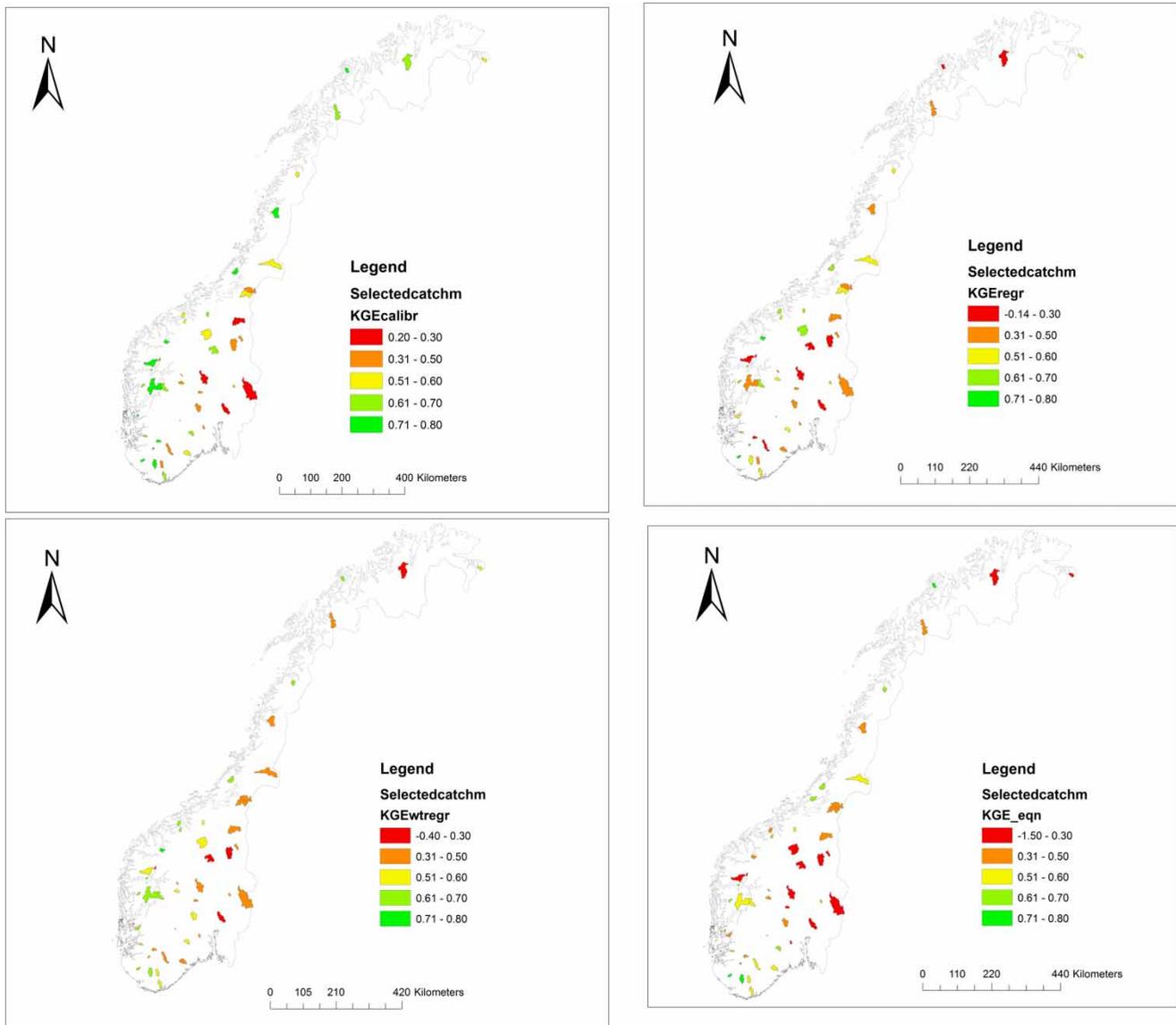
Figure 7 │ Results for median KGE values from simulation using the average parameters values from calibration, equations (Equations (1)–(3)), weighted regression (Equations (7)–(9)) and regression (Equations (4)–(6)).

slow response rate (Table 4). An opposite correlation is observed for the coefficient of variation for K_1; the values are higher for slow responding catchments that have a larger A, a lower $DrInd$ and a low Hl. There is a much higher degree of spatial variability in the rainfall intensity and soil saturation and this variability presumably leads to larger model errors in these catchments. The higher uncertainty in the K_1 values can be explained by the inclusion of catchments with relatively large areas, such as the largest in the data set, for which the PQRUT

model is generally not used. This catchment was also an outlier for the regression equations for K_1. The time of concentration increases with catchment size and in more advanced hydrological models this is accommodated by various storage mechanisms. For this reason, the PQRUT model simulations based on a common set of parameters for all catchments performs well for some events, but very poorly for certain individual events (Figure 8).

In order to reduce the influence of the catchments where the precision of the estimation of the mean parameter

Table 7 | Mean and median KGE values for simulated streamflow

	Median KGE	Mean KGE
Calibration	0.64	0.58
Weighted regression (Equations (7)–(9))	0.51	0.48
Weighted regression LOOCV (leave-one-out cross-validation)	0.49	0.47
Weighted regression LOOCV, K_1 and T based on calibrated K_2	0.56	0.5
Regression (Equations (4)–(6))	0.49	0.45
Regression LOOCV	0.46	0.43
Regression LOOCV, K_1 and T based on calibrated K_2	0.48	0.44
Equations (Equations (1)–(3))	0.46	0.33

values is low, the weighted regression method was used. The catchments that had the lowest weights in the regression for K_1 were medium-sized catchments with spring flood regimes, which means that there was a significant snowmelt contribution during the highest flow events. These events were not selected for calibration because the snowmelt was higher than 20%, and an hourly model for disaggregating the daily snowmelt was not available. Therefore, relatively insignificant peak flows often occurring during the summer and autumn months were used for the calibration, and this contributes to the higher uncertainty in the K_1 values for these catchments. For some of these catchments, the values of K_1 may be related to different generation processes for different flood hydrographs, and as a result, the calibrated K_1 values are correlated with different catchment descriptors. The catchments that had the highest standard error for K_2 and T had an area of less than 100 km^2 (excepting one with an area of 267 km^2) and different seasons of highest floods. The uncertainty reflects the small size of the catchment and, in these cases, precipitation data of higher temporal resolution may be required
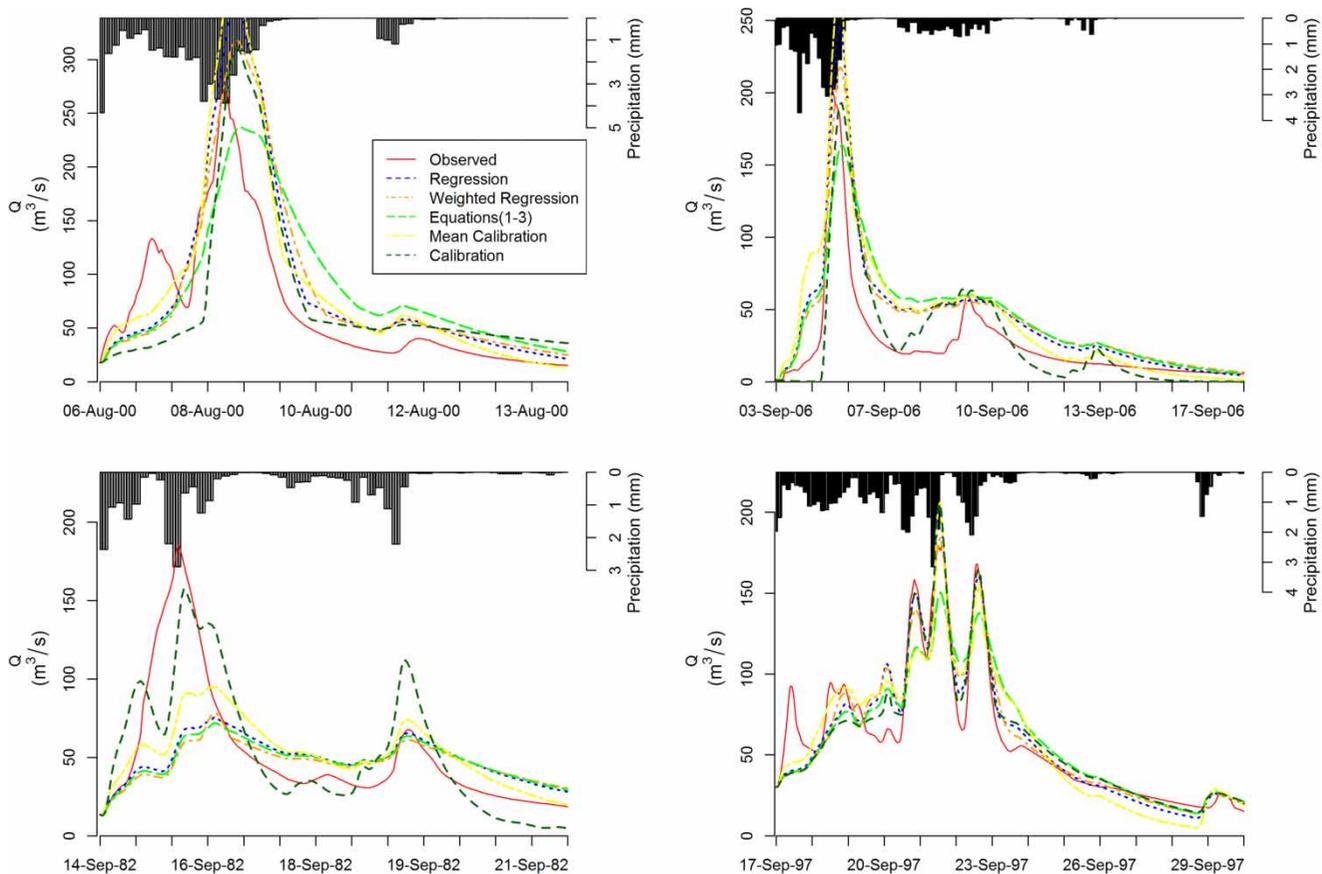

Figure 8 | Observed and simulated hydrographs, using parameters, estimated from: regression, weighted regression, Equations (1)–(3), mean parameter set from calibration and calibration for each event, for selected flood events for one catchment of area 480 km^2. The legend shown in the upper left figure applies to all diagrams.

for accurate flow simulation instead of the data used here, which are 3-hourly values disaggregated from daily values.

CONCLUSION

In this study, the three parameters of PQRUT, an event-based model used for extreme flood estimation, have been regionalised using calibration based on disaggregated precipitation data. There is an improvement in the performance of this regionalisation as compared with the equations developed by Andersen *et al.* (1983). In addition, the weighted regression method increased the median KGE, as compared with OLS regression, although there is still a large degree of uncertainty in the regression model which appears to be associated with the interdependency of the parameter values. The use of a 1-hour time step allowed for the accurate simulation of peak flows in most catchments, but higher temporal resolution appears to be needed for calibration of the smallest catchments, i.e., less than 100 km^2. The use of disaggregated seNorge precipitation data, using the HIRLAM model, may have increased the uncertainty in the calibration. The results from the calibration may be improved if, for example, spatially interpolated precipitation data with an hourly time step were available. The difficulty of assigning appropriate initial conditions at the start of the simulation has also contributed to uncertainty in the calibration. The use of numerous peak events for each catchment in the calibration procedure allowed for an investigation of parameter uncertainty. A measure of the uncertainty as weights was introduced into the weighted regression in order to reduce the influence of catchments with highly variable parameter values.

The results showed that there is no significant spatial autocorrelation and geographically homogeneous regions could not be delineated, except for the catchments in southeast Norway. For this reason, methods based on spatial proximity such as kriging could not be used. The inclusion of a higher number of catchments is needed to determine whether there is spatial autocorrelation of the parameters, and if more catchments are included in southeast Norway, regression equations that are more robust can be developed for this subregion. A consideration can be made to also

regionalise the values of the parameter lp which accounts for the soil moisture deficit in this area.

REFERENCES

Andersen, J., Hjukse, T., Roald, L. & Sælthun, N. 1983 *Hydrologisk Modell for Flomberegninger.* Report No. 2 1983, NVE, Oslo (in Norwegian).

Blazkova, S. & Beven, K. 2004 Flood frequency estimation by continuous simulation of subcatchment rainfalls and discharges with the aim of improving dam safety assessment in a large basin in the Czech Republic. *J. Hydrol.* **292**, 153–172.

Blöschl, G. & Sivapalan, M. 1995 Scale issues in hydrological modelling: a review. *Hydrol. Process.* **9**, 251–290.

Burn, D. H. & Boorman, D. B. 1992 *Catchment classification applied to the estimation of hydrological parameters at ungauged catchments.* Institute of Hydrology Report No. 118, Wallingford, UK.

Filzmoser, P., Maronna, R. & Werner, M. 2008 Outlier identification in high dimensions. *Comput. Stat. Data Anal.* **52** (3), 1694–1711.

Fleig, A., Andreassen, L., Barford, E., Haga, J., Haugen, L., Hisdal, H., Melvold, K. & Saloranta, T. 2013 *Norwegian Hydrological Reference Dataset for Climate Change Studies.* Report No. 2 2013, NVE, Oslo, Norway.

Fox, J. & Weisberg, S. 2011 *An R Companion to Applied Regression.* Sage, Thousand Oaks, CA, USA.

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A., Merz, R. & Blöschl, G. 2015 Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrol. Sci. J.* **60** (6), 968–984.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modeling. *J. Hydrol.* **377** (1–2), 80–91.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. & Cudennec, C. 2013 A decade of predictions in ungauged basins (PUB) – a review. *Hydrol. Sci. J.* **58** (6), 1198–1255.

Kaufman, L. & Rousseeuw, P. 1990 *Finding Groups in Data: Introduction to Cluster Analysis.* Wiley, NY, USA.

Kottegoda, N. T. & Rosso, R. 1997 *Statistic, Probability, and Reliability for Civil and Environmental Engineers.* McGraw-Hill, NY, USA.

Ladson, A. R., Brown, R., Neal, B. & Nathan, R. 2013 A standard approach to baseflow separation using the Lyne and Hollick filter. *Aust. J. Water Resour.* **17** (1), 25–34.

Lawrence, D., Haddeland, I. & Langsholt, E. 2009 *Calibration of HBV hydrological models using PEST parameter estimation*. Report No. 1 2009, NVE, Oslo, Norway.

Lawrence, D., Paquet, E., Gailhard, J. & Fleig, A. K. 2014 Stochastic semi-continuous simulation for extreme flood estimation in catchments with combined rainfall–snowmelt flood regimes. *Nat. Hazards Earth Syst. Sci.* **14**, 1283–1298.

Lee, H., McIntyre, N. R., Wheater, H. S. & Young, A. R. 2006 Predicting runoff in ungauged UK catchments. *Proc. Inst. Civ. Eng. Water Manage.* **159**, 129–138.

Loukas, A. & Kjeldsen, T. (eds) 2013 Advanced methods for flood estimation in a variable and changing environment. *Nat. Hazards Earth Syst. Sci.* 14.

Lyne, V. & Hollick, M. 1979 Stochastic time-variable rainfall-runoff modelling. In: *Proceedings Hydrology and Water Resources Symposium*. Institution of Engineers, Perth, Australia, pp. 89–92.

McIntyre, N., Lee, H., Wheater, H., Young, A. & Wagener, T. 2005 Ensemble predictions of runoff in ungauged catchments. *Water Resour. Res.* **41**, W12434.

Merz, R. & Blöschl, G. 2004 Regionalisation of catchment model parameters. *J. Hydrol.* **287**, 95–123.

Merz, R. & Blöschl, G. 2005 Flood frequency regionalisation – spatial proximity vs. Catchment Attributes. *J. Hydrol.* **302**, 283–306.

Mitchell, A. 2005 *The ESRI Guide to GIS Analysis, Volume 2: Spatial Measurements and Statistics*, 1st edn. ESRI Press, Redlands, CA, USA.

Mohr, M. 2008 *New Routines for Gridding of Temperature and Precipitation Observations for 'seNorge.no'*. Note no. 08/2008 Met.no, Oslo, Norway.

Paquet, E., Garavaglia, F., Garcon, R. & Gailhard, J. 2013 The SCHADEX method: a semi-continuous rainfall-runoff simulation for extreme flood estimation. *J. Hydrol.* **495**, 23–37.

Razavi, T. & Coulibaly, P. 2013 Streamflow prediction in ungauged basins: review of regionalisation methods. *J. Hydrol. Eng.* **18** (8), 958–975.

Seibert, J. 1999 Regionalisation of parameters for a conceptual rainfall runoff model. *Agric. Forest Meteorol.* **98–99**, 279–293.

Serinaldi, F. & Grimaldi, S. 2011 Synthetic design hydrographs based on distribution functions with finite support. *J. Hydrol. Eng.* **16** (5), 434–446.

Skaugen, T., Peerebom, I. O. & Nilsson, A. 2014 Use of a parsimonious rainfall–run-off model for predicting hydrological response in ungauged basins. *Hydrol. Process.* **29**, 1999–2013.

Tolson, B. A. & Shoemaker, A. 2007 Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resour. Res.* **43**, W01413.

Viviroli, D., Mittelbach, H., Gurtz, J. & Weingartner, R. 2009 Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: parameter regionalisation and flood estimation results. *J. Hydrol.* **377**, 208–225.

Vormoor, K. & Skaugen, T. 2013 Temporal disaggregation of daily temperature and precipitation grid data for Norway. *J. Hydrometeor.* **14**, 989–999.

Vormoor, K., Skaugen, T., Langsholt, E., Diekkrüger, B. & Skøien, J. 2011 Geostatistical regionalization of daily runoff forecasts in Norway. *Int. J. River Basin Manage.* **9** (1), 3–15.

Wagener, T., Wheater, H. S. & Gupta, H. V. 2004 *Rainfall-Runoff Modelling in Gauged and Ungauged Catchments*. Imperial College Press, London, UK.

Wilson, D., Fleig, A. K., Lawrence, D., Hisdal, H., Pettersson, L. & Holmqvist, E. 2011 *A review of NVE's flood frequency estimation procedures*. Report No. 9 2011, NVE, Oslo, Norway.

Yeh, K. C., Yang, J. C. & Tung, Y.-K. 1997 Regionalisation of unit hydrograph parameters: 2. Uncertainty analysis. *Stoch. Hydrol. Hydraul.* **11**, 173–192.