



# A method for low-flow estimation at ungauged sites: a case study in Wallonia (Belgium)

M. Grandry, S. Gailliez, C. Sohier, A. Verstraete, and A. Degré

Univ. Liège – Gembloux Agro-Bio Tech., Soil–Water Systems, Passage des Déportés, 2, 5030 Gembloux, Belgium

Correspondence to: A. Degré (aurore.degre@ulg.ac.be)

Received: 2 October 2012 – Published in Hydrol. Earth Syst. Sci. Discuss.: 11 October 2012

Revised: 8 March 2013 – Accepted: 14 March 2013 – Published: 8 April 2013

**Abstract.** Well-integrated water management can notably require estimating low flows at any point of a river. Depending on the management practice, it can be needed for various return periods. This is seldom addressed in the literature. This paper shows the development of a full analysis chain including quality analysis of gauging stations, low-flow frequency analysis, and building of a global model to assess low-flow indices on the basis of catchment physical parameters.

The most common distributions that fit low-flow data in Wallonia were two-parameter lognormal and gamma. The recession coefficient and percolation were the most explanatory variables, regardless of the return period. The determination coefficients of the models ranged from 0.51 to 0.67 for calibration and from 0.61 to 0.80 for validation. The regression coefficients were found to be linked to the return period. This was used to design a complete equation that gives the low-flow index based on physical parameters and the desired return period (in a 5 to 50 yr range).

The interest of regionalisation and the development of regional models are also discussed. Four homogeneous regions are identified, but to date the global model remains more robust due to the limited number of 20-yr-long gauging stations. This should be reconsidered in the future when enough data will be available.

important during low flows, some conflicts between the different water uses can arise, especially between instream water use and water abstraction demand (Hébert et al., 2003). Water managers therefore need to be able to quantify low flows at any point of a river, in magnitude as well as in frequency.

Low flows can have different meanings depending on the definitions of authors. In this study, low flows are considered as the lowest discharge values observed in a river, which usually occur between May and November in Wallonia (Belgium). The index chosen to characterise low flows is MAM7 which stands for mean annual minimum flow on a 7-day average basis.

In gauged catchments, the value of MAM7 is calculated on a yearly basis. Then, a distribution can be fitted to the data in order to find a relationship between low flows and probabilities of non-exceedance ( $1/T$ ). The distribution that best fits is chosen according to statistically and graphically based tests, and its parameters can be calculated by different methods (Smakhtin, 2001). Matalas (1963), Joseph (1970), Condie and Nix (1975), Tasker (1987), Leppärjärvi (1989) and Yue and Pilon (2005), amongst many others, have compared various distributions and methods to estimate parameters. The best-adjusted distribution is usually different according to the study region and the low-flow index (Abi-Zeid and Bobée, 1999). In other methods, the understanding of the generating processes of low flows contributes to the choice of distribution (Pacheco et al., 2006).

Low-flow calculation and frequency analysis are easy to handle for long-time gauged catchments. For ungauged catchments, however, low-flow index has to be inferred using neighbouring gauged catchment data.

## 1 Introduction

It is now recognised that river low flows can lead to severe consequences in water quality and river ecological status (Whitehead et al., 2009). Navigation and power supply sectors can also be affected by low flows (Middelkoop et al., 2001). Furthermore, as pressures on rivers become more

Regarding the estimation of low flows in ungauged river catchments, Smakhtin (2001) reviewed all possible techniques but cited regional regression approach as the most widely used. The other methods use time-series simulation, regional prediction curves, or spatial interpolation. Regional regression consists in delineating hydrologically homogeneous regions based on catchment characteristics and developing, for each region, a regression model relating the low-flow index to these characteristics. As this approach is based on physical parameters of catchments, it allows a better understanding of low flows from a physical point of view, in addition to estimating them. This technique has been employed by numerous scientists, e.g. Nathan and McMahon (1992) for south-eastern Australia, Laaha and Blöschl (2006) for Austria, Vezza et al. (2010) for north-western Italy, and Tsakiris et al. (2011) for the state of Massachusetts in the USA. However, only a few of them included the return period in their analysis.

The temporal and spatial components of low-flow hydrology such as frequency and regional distribution, respectively, are closely related according to Smakhtin (2001), but articles that deal with both at the same time are rare. Some of them have aimed at assessing low flow for fixed return periods. Kim and Lee (2010) used Bayesian multiple regression analysis for regional low-flow frequency in south-east Korea and utilised the developed model to predict 7-day minimum flow at ungauged sites for a fixed 10 yr return period. Comparing to conventional techniques based on a  $t$  distribution, the Bayesian analysis resulted in lower confidence intervals for parameter estimates. After a low-flow frequency analysis comparing nine distributions, Ouarda and Shu (2009) developed models to estimate low flows in Quebec, employing single and ensemble artificial neural networks, for three different return periods and taking into account seven physiographical and meteorological variables. In Virginia, Hayes (1991) used the Pearson type III distribution for frequency analysis. Then, he developed regression models to estimate the annual minimum average 7-consecutive-day discharge from basin and climate characteristics at ungauged sites for two different return periods. Saravi et al. (2010) applied the regional regression approach for frequency analysis of annual peak maximum series of flood flows and also used the models to estimate flood quantiles at ungauged sites for several return periods. They considered seven climatic and physical catchment characteristics.

The only study which aimed at assessing a low-flow index for a range of return period is that of Chen et al. (2006). They performed a frequency analysis of low flows comparing five distributions and using  $L$  moments for the Dongjiang basin in South China. The quantile function of the 3-parameter log-normal distribution, which was the most appropriate distribution for the study data set, was then integrated into this regression model in order to establish the formula of the mean 7-day low flow for any return period. But they used a limited

data set of 14 sites and used the catchment area as the only characteristic to assess the mean 7-day low flow.

To date, there is no mean to estimate low-flow indices in ungauged rivers for a range of return periods, on the basis of a set of meaningful physical parameters. Hence, in the case of droughts, different measures to maintain a minimum flow in rivers could need this information (e.g. “Plan d’Action Sécheresse” in France, drought plans in the UK).

The aim of this study is to propose a complete method to build a model that can estimate low flows for any return period in ungauged rivers. We propose a full analysis chain: a selection of quality gauging stations and a low-flow frequency analysis with comparison of different distributions are first carried out, followed by the development of a regression model that uses physical parameters of the catchments. We aim at obtaining a formula that can be used for any return period between 5 and 50 yr. To do so, we propose a new approach which evaluates the relationships between regression coefficients and the return period. The regionalisation of low flows is also tested in order to improve the model performances.

## 2 Material and methods

### 2.1 Study area

The Walloon Region of Belgium covers an area of 16 844 km<sup>2</sup>. The two main catchments crossing this region are the Meuse (70 % of the area of Wallonia) and the Scheldt (20 % of Wallonia) catchments. Wallonia is characterised by a high number of small basins (70 % of gauged catchments are smaller than 200 km<sup>2</sup>).

### 2.2 Choice of low-flow index

The mean annual minimum of 7-day average flows (MAM7), which is one of the most widely used indices, was selected for this study. The main advantages of this parameter are that it eliminates day-to-day variations and allows analyses to be less sensitive to measurement errors. Moreover, 7-day low flows are not very different from 1-day low flows (Smakhtin, 2001). Averaging over some days also allows smoothing out some human influences on flows such as variation of hourly flows due to hydropeaking and little abstraction from farmers.

### 2.3 Selection of gauging stations

More than 240 gauging stations have been installed in Wallonia during the last 40 yr. From those, we selected the stations that fulfilled the following criteria:

- minimum of 20 yr of data (Laaha and Blöschl, 2005);
- homogeneous data: homogeneity tests, through tests of equality of means, were carried out for stations that

have been moved or for which the reference level has changed following the replacement of measurement devices;

- no human influence on flows (dams, abstraction, etc.);
- no data extrapolated from another station in the same catchment, to avoid errors in flow values and heterogeneity in data;
- MAM7 over  $5 \text{ L s}^{-1}$  because rating curves are not precise enough below this value; and
- no aquatic vegetation in summer, to avoid overestimation of low flows.

Finally, 59 gauging stations were selected. The main rejection reason was a too short chronicle.

## 2.4 Frequency analysis

The data used for this analysis are the annual minimum 7-day average flow series that we abbreviated to AM7 and correspond to  $7Q$  or  $Q_7$ , which is generally used in the USA for return periods of 2 and 10 yr ( $7Q_2$  and  $7Q_{10}$  or  $Q_{7,2}$  and  $Q_{7,10}$ ) (Smakhtin, 2001; Hayes, 1991; Vogel and Kroll, 1989).

We performed a frequency analysis in order to predict AM7 for return periods ( $T$ ) of 5, 10, 20 and 50 yr ( $AM7_T$ ). As the length of available data is quite short (maximum of 45 yr),  $AM7_T$  cannot be predicted with accuracy for a return period higher than 50 yr.

For each station, 2-parameter lognormal, 2-parameter Weibull, gamma, Fréchet, 3-parameter lognormal and 3-parameter Pearson distributions were tested. These six distributions were chosen because they are the most often used (Matalas, 1963; Joseph, 1970; Condie and Nix, 1975; Xanthoulis, 1985; Yue and Pilon, 2005; Chen et al., 2006; Modarres, 2008).

Parameters of all distribution laws were estimated using the maximum likelihood procedure. Indeed, this method provides asymptotically minimum variance estimates, is adapted to all distributions and to low flows, and has given good results in other studies (Joseph, 1970; Condie and Nix, 1975; Landwehr et al., 1979; Leppärjärvi, 1989; Nathan and McMahon, 1990).

These distributions, except Fréchet, were ordered by increasing posterior probability, and decreasing Akaike's information criterion (AIC) and Bayesian information criterion (BIC). Posterior probabilities were calculated from prior probabilities and Bayes factors. These factors were approximated via Schwarz's method (Schwarz, 1978). As no a priori information on the suitability of each law was available, prior probabilities were considered equal for all distribution laws. AIC and BIC both take into account the likelihood function and the number of parameters, but BIC also considers the

size of the sample. Therefore, 2-parameter distributions were favoured by this ranking.

The statistical  $\chi^2$  test was also carried out to check the adjustment of each distribution to the sample.

Finally, the Fréchet distribution and the three best distributions for which the  $\chi^2$  test hypothesis was accepted were compared graphically. The graph showed observed data (AM7) in function of probabilities of non-exceedance as well as frequency curves for the four distributions. If two different distributions gave the same fit, the simpler one with less parameters was selected (Miquel, 1984). Once the best-fitted distribution was chosen,  $AM7_T$  was estimated for the four return periods, along with their 95 % confidence interval (CI).

The selection of the best distribution, except for the Fréchet distribution, was performed using HYFRAN (HYdrological FREquency ANalysis) software which was created, for the purpose of fitting statistical laws, by B. Bobée from the National Institute for Scientific Research – Water, Soil, Environment Centre (University of Quebec) (El Adlouni et al., 2008).

## 2.5 Development of a global regression model

It is acknowledged that catchments which have similar physical and climatic features have similar hydrological responses (Smakhtin, 2001). Firstly, we developed a regression model for the whole study area to estimate AM7 from catchment characteristics. Secondly, we tested if homogeneous regions can be identified in the area, using regionalisation techniques. In these regions, we developed regional models and compared their performance with the global model.

In this study, climatic and physical catchment characteristics were described by the 25 variables presented in Table 1. “Summer” refers to the July to September period and “winter” refers to the October to April period. Position data (Alt, X and Y) were measured with a GPS or by levelling. Catchment boundaries were defined using a digital terrain model (DTM). Catchment features (A, DD, SIs, Ls and Ss) were derived from GIS maps. The hydrological type of soil describes the infiltration rate (high for A to very low for D) and drainage (excellent for A to very bad for D) of soils (Demarcin et al., 2011). Meteorological data (AP, SP, WP, ST) were interpolated for each catchment by the hydrological model EPICgrid (Sohier et al., 2009) from meteorological data measured at some locations in Wallonia, by means of the Thiessen polygon method. PET was also computed by the model, using the Penman equation. This model is actually able to simulate the atmosphere–soil–plant continuum, using soil, geology, land use, agricultural practices, topography (DTM) and meteorological data. Percolation is defined as the quantity of rainfall that reaches deeper soil layers. It was estimated for each catchment using a “capacitive” approach: each soil layer is considered as a tank that empties

**Table 1.** The 25 physical and climatic variables considered in this study to describe catchments, along with their abbreviation and unit. “Summer” refers to the July to September period and “winter” refers to the October to April period.

Variable	Abbreviation	Unit
Altitude of the gauging station	Alt	m
Map coordinates of the station	X	m
	Y	m
Catchment area	A	km <sup>2</sup>
Drainage density	DD	km km <sup>-2</sup>
10th percentile of the slope	Sl <sub>10</sub>	%
50th percentile of the slope	Sl <sub>50</sub>	%
90th percentile of the slope	Sl <sub>90</sub>	%
Area percentage of urban lands	L <sub>u</sub>	%
Area percentage of forests	L <sub>f</sub>	%
Area percentage of arable lands	L <sub>a</sub>	%
Area percentage of permanent crops	L <sub>p</sub>	%
Area percentage of grasslands	L <sub>g</sub>	%
Area percentage of soils of hydrological group A	S <sub>B</sub>	%
Area percentage of soils of hydrological group B		
Area percentage of soils of hydrological group C	S <sub>C</sub>	%
Area percentage of soils of hydrological group D	S <sub>D</sub>	%
Area percentage of soils that were not mapped	S <sub>NM</sub>	%
Annual precipitation	AP	mm
Summer precipitation	SP	mm
Winter precipitation	WP	mm
Summer temperature	ST	°C
Potential evapotranspiration	PET	mm
Percolation	Pe	mm
Recession coefficient	RC	day <sup>-1</sup>

when the layer water content is greater than field capacity (Sohier et al., 2009).

Recession is the part of stream flow in which discharge depletes gradually and there is no rainfall or human influence (Dacharry, 1997; Tallaksen, 1995). The recession coefficient is the parameter of the exponential model that describes the recession process. It was calculated for each catchment, using the method developed by Lang and Gille (2006). Recession periods were first defined according to flow and precipitation thresholds, and by removing overland flow influence. Thresholds were adapted to Wallonia. A mean or master recession curve was then constructed using a technique based on the correlation method. The recession coefficient is the parameter of this exponential curve.

Since catchment area was highly correlated to AM7<sub>T</sub> (correlation coefficient of 0.72 for  $T = 5$  yr), we used specific flows (AM7<sub>T</sub> divided by the area) as the dependent variable. The explanatory variables were the 24 other climatic and physical catchment characteristics.

Out of these 24 variables, it was necessary to select a few of them that were not correlated to each other and that could

explain the most variability of specific AM7<sub>T</sub>. For this purpose, we used three different methods: stepwise, maximum  $R^2$  improvement and adjusted  $R^2$  selection. Regression coefficients were estimated using the ordinary least squares technique. As the second method gives the best model for each number of variables ( $p$ ), the one with the Mallows coefficient ( $C_p$ ) close to  $p + 1$  was selected (Mallows, 1995). For the two last methods, non-significant variables ( $p$  value > 0.05) were removed one by one from the model. Also, variables with a variance inflation factor (VIF) above 10 were deleted because this indicates a multicollinearity problem (Confais and Le Guen, 2006; Vezza et al., 2010).

The presence of outliers could be detected by a Cook’s distance (Cook’s  $D$ ) above 1 (Confais and Le Guen, 2006; Laaha and Blöschl, 2006).

The normality and the equality of variance of residuals were evaluated graphically:  $Q-Q$  plot for normality and residuals-predicted AM7<sub>T</sub> plot for the equality of variance (residuals must be around 0) (Confais and Le Guen, 2006; Vezza et al., 2010).

The model was calibrated using the variables related to the catchments of the 59 selected gauging stations. The validation of the models was performed using another data set associated with 19 stations which met all criteria, but their record length was between 15 and 20 yr. The validation sample was representative of the study area as these stations are spread over Wallonia.

In order to compare the performance of the models obtained by the three methods,  $R^2$ , adjusted  $R^2$  and RMSE (root-mean-square error) were calculated for each calibration and validation according to the following formulae (Laaha and Blöschl, 2006; Vezza et al., 2010):

$$R^2 = \frac{\text{var}(Y) - \text{MSE}}{\text{var}(Y)}, \quad (1)$$

$$R^2_{\text{adj}} = 1 - \frac{n-1}{n-p-1}(1-R^2), \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}, \quad (3)$$

in which  $\text{var}(Y)$  is the variance of observed AM7<sub>T</sub>,  $n$  is the number of observations,  $p$  is the number of variables in the model,  $Y_i$  is the observed value of AM7<sub>T</sub> for the observation  $i$ , and  $\hat{Y}_i$  is the predicted value.  $R^2$  gives the part of the variability of AM7<sub>T</sub> explained by the model.  $R^2$  can be adjusted for the number of explanatory variables in the model, which is useful to compare different models. When a variable is added to the model, unlike  $R^2$ , the adjusted  $R^2$  increases only if the new variable has additional predictive capability. RMSE quantifies the difference between observed and predicted AM7<sub>T</sub>.

To evaluate the uncertainty of the model, the prediction interval can be calculated. It represents the confidence interval of predicted values of AM7, i.e. plausible values of AM7 for given values of explanatory variables. It takes into account the errors associated with the calibration of the model and with the data used to calculate the predicted value. The  $100(1-\alpha)$  percent confidence interval for the predicted value  $\hat{Y}_p$  is obtained as follows (Dagnélie, 2006):

$$\hat{Y}_p \pm t_{1-\alpha/2} \sqrt{s^2(1 + X'_p(\mathbf{X}\mathbf{X})^{-1}X_p)}, \quad (4)$$

in which

- $t_{1-\alpha/2}$  is the tabulated  $t$  quantile from the Student distribution with degrees of freedom equal to  $n-p-1$ ,
- $s$  is the estimate of the residual variance (the mean squared error (MSE)),
- $\mathbf{X}$  is the matrix of the observations used to calibrate the model, and
- $X_p$  is the vector of observations used to calculate the predicted value.

All statistical analyses were performed using SAS (statistical analysis system) software.

The last step was to evaluate the relationships between regression coefficients and the return period in order to insert the return period as a variable in the equations.

## 2.6 Low-flow regionalisation

As mentioned above, we tested the possibility of delineating homogeneous regions within the study area. Catchment and climate variables were standardised and the ones describing land use and soils were weighted (divided by the square root of the number of variables per characteristic).

Homogeneous regions were obtained by performing a cluster analysis with the 25 variables for the 59 catchments. The clustering method used was the agglomerative hierarchical clustering which merges two small groups into a bigger one. The starting clusters were 59 groups of one observation each.

Since all variables are quantitative, we chose Ward's algorithm (Ward, 1963) as the merging strategy. Moreover, this method has often been used in low-flow regionalisation (Laaha and Blöschl, 2006; Vezza et al., 2010). Each merger was then carried out in order to have the smallest difference in  $R^2$  between two groupings. The determination coefficient  $R^2$  represents the proportion of information kept after the fusion of clusters. The fusions were stopped before this difference in  $R^2$  became large.

To interpret the results of clustering and characterise the regions, we carried out a principal component analysis (PCA) which helps understand the main differences between groups. The mean and standard deviation of all variables for

each group were also calculated and boxplots were drawn. This allowed locating groups in the plane of variables and comparing groups according to these 25 variables.

## 3 Results

### 3.1 Frequency analysis

For each of the 59 stations, the distribution that best fitted the data was chosen amongst 2-parameter lognormal, 2-parameter Weibull, gamma, Fréchet, 3-parameter lognormal and 3-parameter Pearson distributions. Two-parameter lognormal and gamma are the most common laws in Wallonia. No relationship was found between the type of distribution selected and the length of data, catchment area or the spatial location of the catchment.

### 3.2 Development of a global regression model

Applicability conditions were checked and the logarithm of AM7<sub>T</sub> was chosen in order to improve the normality of residuals. In addition, Laaha and Blöschl (2006) proposed to use a logarithmic transformation when outliers increase with observed flow, which occurred in our case. Finally, this transformation allows avoiding negative estimates of AM7<sub>T</sub>.

The Cook distance allowed us to detect three outliers. These observations corresponded to higher specific AM7. Removing them from the data set would have improved the fitting of the model in calibration, but this was not a sufficient reason. They were therefore kept, but one should note that the model is better calibrated within a range of specific AM7.

For each regression method (stepwise, maximum  $R^2$  improvement and adjusted  $R^2$  selection), the same variables were selected for all return periods, except the 50 yr return period for the two last methods. Maximum  $R^2$  improvement and adjusted  $R^2$  selection actually resulted in the same final equations, except for the 50 yr return period. Table 2 gives the regression coefficients for each method and return period. The general equation is

$$\text{AM7}_T = \text{AREA} \times 10^{\text{constant} + ax_1 + bx_2 + \dots + zx_p}. \quad (5)$$

Equation (6) is an example of the equation for AM7<sub>T5</sub> obtained by the stepwise method:

$$\text{AM7}_{T5} = \text{AREA} \times 10^{-2.7851 + 0.0017 Pe - 13.4274 RC}. \quad (6)$$

Table 3 presents, for each method, the values of  $R^2$ , adjusted  $R^2$  and RMSE for the calibration and the validation of the models.

For calibration, the performance of the two models is similar except for the 50 yr return period for which the models obtained by the maximum  $R^2$  improvement and adjusted  $R^2$  selection perform better (higher adjusted  $R^2$  and lower RMSE). However, the validation is clearly better with the

**Table 2.** Regression coefficients of the global model by method of regression and return period. The general equation is  $AM7_T = AREA \times 10^{\text{constant} + ax_1 + bx_2 + \dots + zx_p}$ .  $L_g$  is the area percentage of grasslands [%],  $S_A$  the area percentage of soils of hydrological group A [%], WP winter precipitation [mm], RC recession coefficient [ $\text{day}^{-1}$ ], Pe percolation [mm],  $Sl_{90}$  the 90th percentile of slope [%],  $S_B$  the area percentage of soils of hydrological group B [%], and  $S_C$  the area percentage of soils of hydrological group C [%].

Method	Return period	Constant	$L_g$	$S_A$	WP	RC	Pe	$Sl_{90}$	$S_B$	$S_C$
Max. $R^2$ improv.	5	-3.8935	0.0080	0.0098	0.0017	-19.8274	-	-	-	-
$R^2_{\text{adj}}$ selection	10	-3.9741	0.0089	0.0104	0.0017	-21.2826	-	-	-	-
	20	-4.0639	0.0098	0.0110	0.0017	-22.4033	-	-	-	-
Max. $R^2$ improv.	50	-4.4764	-	0.0154	0.0016	-20.4359	-	0.0348	-	-
$R^2_{\text{adj}}$ selection	50	-3.1899	-	-	0.0017	-19.1631	-	0.0409	-0.0148	-0.0183
Stepwise	5	-2.7851	-	-	-	-13.4274	0.0017	-	-	-
	10	-2.8396	-	-	-	-15.5904	0.0016	-	-	-
	20	-2.8939	-	-	-	-17.3211	0.0016	-	-	-
	50	-2.9808	-	-	-	-19.0192	0.0016	-	-	-

model obtained by the stepwise method for all return periods. Since the aim of this study is to be able to estimate  $AM7_T$  in ungauged catchments, we continued the analyses with the models obtained by stepwise. According to Laaha and Blöschl (2007), the stepwise method maximises the robustness and the predictive performance of the model, and minimises collinearity between variables. Our results confirm the lower collinearity when using the stepwise method. If we consider the logarithms of specific  $AM7$ , VIF is 1.09 for the stepwise method, while VIFs range from 1.4 to 3.21 for the maximum  $R^2$  improvement method and from 1.4 to 2.68 for the adjusted  $R^2$  selection method.

For the stepwise model, the  $R^2$  and RMSE of calibration decrease when the return period increases, which means that the part of the variance of  $AM7_T$  explained by the model and residuals both decrease. This seems contradictory but can be explained by the diminution in the variance of observed  $AM7_T$  when  $T$  increases, this diminution being relatively bigger than the reduction in MSE. From a validation point of view,  $R^2$  is quite high but diminishes also when  $T$  becomes larger. RMSE decreases as well but is in the same range of values as the RMSE of calibration. Therefore, the variance of observed  $AM7_T$  of validation stations is higher. In conclusion, the model performs generally well and even better for predicting since the part of the variance of  $AM7_T$  explained by the model is higher in validation. This performance is detailed in the following paragraphs.

Residuals increase linearly with observed specific  $AM7_T$ , as can be observed in Fig. 1 for T5. Therefore, the models overestimate low values of specific  $AM7_T$  (especially under  $10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T5 and T10, and under  $5 \times 10^{-4} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T20 and T50), and underestimate higher values of specific  $AM7_T$  (especially over  $3 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T5 and T10, and over  $2 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T20 and T50). This problem of lack of calibration for very high and low values of specific

$AM7_T$  is due to the small number of observations for this range of specific  $AM7_T$ , especially for specific  $AM7_T$  over  $4 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T5, T10 and T20, and over  $3 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for T50.

The limits of prediction intervals can be calculated using Eq. (4) for the logarithm of specific  $AM7_T$ . In our case, for a risk  $\alpha$  of 5 % and 56 degrees of freedom,  $t$  equals 2.003. The inverse matrix  $(\mathbf{X}'\mathbf{X})^{-1}$  associated with our observations is

$$\begin{Bmatrix} 0.15907243 & -0.00059585 & -3.70977259 \\ -0.00059585 & 0.00000394 & 0.00682459 \\ -3.70977259 & 0.00682459 & 149.582607 \end{Bmatrix}.$$

The sensibility of the model can also be studied. For a constant value of RC, the specific  $AM7_T5$  increases by 3.9 % when Pe increases by 10 mm. For a constant value of Pe, the logarithm of specific  $AM7_T5$  decreases by 6 % when RC increases by  $0.002 \text{ day}^{-1}$ .

Plotting the constant and regression coefficients of the models against the return period, Fig. 2 shows that the constant and regression coefficients are linked to the return period. We adjusted a logarithmic relationship in order to calculate  $AM7_T$  for any return period  $T$  between 5 and 50 yr with this formula:

$$AM7_T = \text{Area} \times 10^{-2.6457 - 0.0847 \ln T + 0.0017 \text{Pe} - 9.8077 \text{RC} - 4 \times 10^{-5} \text{Pe} \ln T - 2.4148 \text{RC} \ln T}. \quad (7)$$

The choice of a logarithmic relationship was made in the context of our results and should be verified in other situations. Compared to  $AM7_T$  predicted with the models for each return period, values estimated by this model are lower by 0.5 to 3 % for 5 yr and 50 yr return periods. For 10 yr and 20 yr return periods, the difference is even lower (between 0.1 and 2 %), and estimated values are generally higher for a 10 yr return period and generally lower for a 20 yr return period. This means that, for 5 yr and 50 yr return periods,

**Table 3.** Comparison of the performance of the models developed using three different regression methods (stepwise, maximum  $R^2$  improvement and adjusted  $R^2$  selection). The performance indices considered for calibration and validation are  $R^2$  (determination coefficient),  $R_{\text{adj}}^2$  (adjusted determination coefficient) and RMSE (root-mean-square error). Maximum  $R^2$  improvement and adjusted  $R^2$  selection resulted in the same final equations, except for the 50 yr return period. The numbers in bold are the highest values of indices for each return period. The stepwise model seems the best option for the purpose of this study.

	T5		T10		T20		T50		
	Stepwise	Max $R^2$ improv. $R_{\text{adj}}^2$ selection	Stepwise	Max $R^2$ improv. $R_{\text{adj}}^2$ selection	Stepwise	Max $R^2$ improv. $R_{\text{adj}}^2$ selection	Stepwise	Max $R^2$ improv.	$R_{\text{adj}}^2$ selection
Calibration $R^2$	0.670	0.700	0.623	0.645	0.578	0.592	0.508	0.794	0.832
Calibration $R_{\text{adj}}^2$	0.658	0.689	0.610	0.632	0.563	0.577	0.490	<b>0.786</b>	<b>0.826</b>
Calibration RMSE ( $\text{m}^3 \text{s}^{-1}$ )	0.201	0.192	0.187	0.181	0.177	0.174	0.170	<b>0.110</b>	<b>0.099</b>
Validation $R^2$	0.796	0.574	0.754	0.510	0.703	0.455	0.608	-0.839	-0.935
Validation $R_{\text{adj}}^2$	<b>0.771</b>	0.453	<b>0.724</b>	0.370	<b>0.666</b>	0.299	<b>0.559</b>	-1.365	-1.679
Validation RMSE ( $\text{m}^3 \text{s}^{-1}$ )	<b>0.205</b>	0.296	<b>0.188</b>	0.265	<b>0.179</b>	0.243	<b>0.177</b>	0.383	0.393

this model slightly underestimates AM7\_T that are already underestimated but improves the estimation of AM7\_T overestimated by the models for each return period. For a 10 yr return period, it is generally the opposite: overestimation of overestimated AM7\_T but improvement of the estimation of underestimated AM7\_T.

### 3.3 Low-flow regionalisation

The cluster analysis resulted in four groups of catchments; the study area could therefore be divided into four homogeneous regions. Catchments in a homogeneous region are contiguous.

The PCA helped to elucidate the differences between groups (Fig. 3a). The first component is highly positively correlated to precipitation, the area percentage of forests, the slopes, the altitude, the area percentage of soils of the hydrological group B and the recession coefficient. It is negatively correlated to the area percentage of arable lands and soils of the hydrological group A, the summer temperature, Y map coordinate, and the area percentage of urban lands. These results have a real physical meaning. The southern region of Wallonia, corresponding to lower Y map coordinates, is characterised by higher altitudes and slopes, while soils of the hydrological group A and arable lands are more present in the northern region of Wallonia where precipitation is smaller and there are also many big city centres.

The second component is positively correlated to percolation and the area percentage of soils of the hydrological group D.

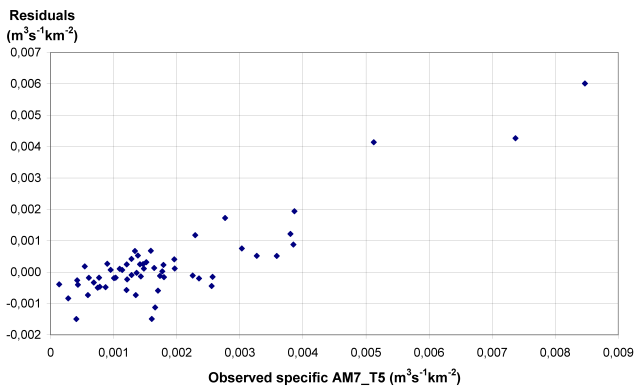
The area percentage of soils of the hydrological group C, soils that were not mapped, permanent crops and grasslands, the area, the drainage density, the X map coordinate and the potential evapotranspiration are variables which are less correlated to the first two components. They contribute therefore less to the grouping of catchments in the regions.

As shown in Fig. 3b, the regions can be distinguished by the climatic and physical features used for the analysis.

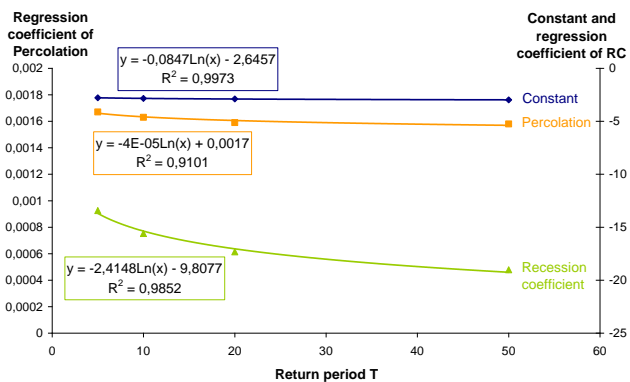
Region 1 is located in the north of Wallonia. It is a region of low altitude (average of 62.4 m) and gentle slopes (average of 3.2 % for the median slope) that receives less precipitation (average of 817 mm), has higher summer temperatures (average of 16.6 °C), is rather agricultural and urban (average of 58.1 % and 10.3 % of the area of catchments covered by arable and urban lands, respectively), and has soils of good infiltration capacity and permeability predominating (average of 44.4 % and 40.7 % of soils of hydrological groups A and B, respectively; 127.5 mm for percolation and 0.0012 for the recession coefficient).

Region 2 has a central spatial location. It is a region characterised mainly by intermediate values of features: medium altitude, steep slopes, average amount of precipitation, average summer temperatures for Belgium, fairly urbanised and agricultural (average of 4.6 % and 22.8 % of the area of catchments covered by urban and arable lands, respectively), rather forested and grassy (average of 30.7 % and 35 % of the area of catchments covered by forests and grasslands, respectively), and soils of moderate infiltration capacity and relatively low permeability predominating (average of 63.4 % and 19.2 % of soils of hydrological groups B and C, respectively; 78.4 mm for percolation and 0.0018 for the recession coefficient).

Region 3 is situated in the south of Wallonia. It is a region of higher altitude (average of 260.8 m) and steep slopes that receives a lot of precipitation (average of 1138 mm), has lower summer temperatures (average of 14.9 °C), is not very urbanised (average of 3.1 % of the area of catchments covered by urban lands), and has soils of relatively good infiltration capacity and moderate permeability predominating (average of 72.6 % and 19.8 % of soils of hydrological groups B and C, respectively; 99.3 mm for percolation and 0.0030 for the recession coefficient), where forests and grasslands prevail (average of 47.4 % and 34.8 % of the area of catchments covered by forests and grasslands, respectively).



**Fig. 1.** Plot of global model residuals against observed specific AM7\_T (example for the return period of 5 yr). Residuals increase linearly with observed specific AM7\_T.

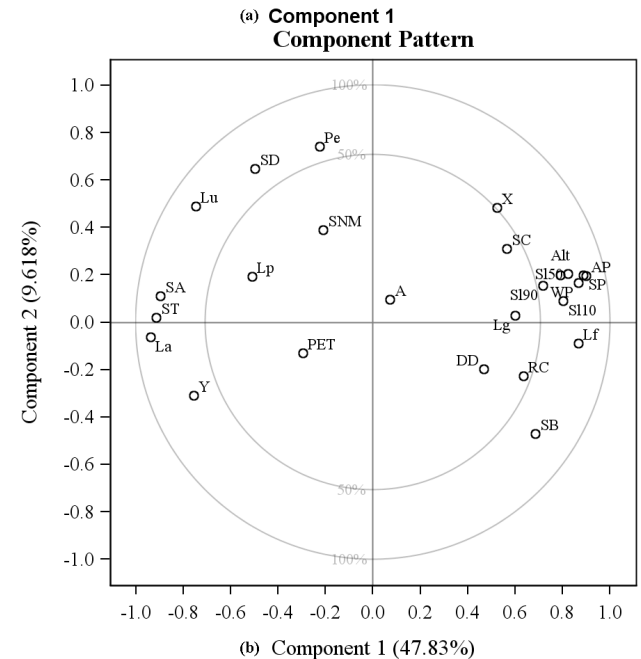
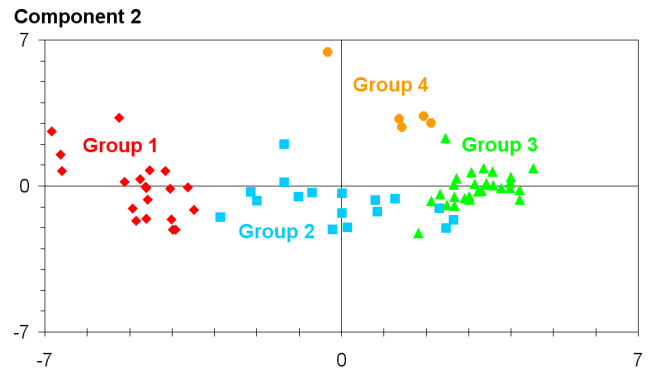


**Fig. 2.** Constant and regression coefficients of the global model plotted against the return period. A logarithmic relationship can be fitted to link the constant and regression coefficients to the return period.

Region 4 is to the south of region 3. It is a region of higher altitude and steep slopes (average of 7.6 % for the median slope) that receives a lot of precipitation (average of 1079 mm), has average summer temperatures, is rather forested and grassy (average of 29.7 % and 36.4 % of the area of catchments covered by forests and grasslands, respectively), and has soils of low infiltration capacity but good permeability predominating (average of 34.3 % and 43 % of soils of hydrological groups B and C, respectively; 218.8 mm for percolation and 0.0012 for the recession coefficient).

**3.4 Development of regional models**

It has been shown by several studies that developing one regression equation per region results in better estimates than a global equation (Smakhtin, 2001; Laaha and Blöschl, 2006; Veza et al., 2010). However, in our case, groups do not contain more than 20 catchments each. Group 4 has only 5 observations, and this can be a problem when checking applicability conditions.



**Fig. 3.** Groups of catchments (a) and correlation circle (b) in the plane of the two first principal components. The two first components of the PCA allow distinguishing the four groups of catchments.

Nevertheless, we developed regional models using the stepwise method for the regions containing enough catchments to calculate statistics and compared their performance with the global model. In region 1, there are 18 catchments for calibration and 6 for validation. Region 3 contains 20 catchments for calibration and 10 for validation.

Applicability conditions were checked and the logarithm of AM7\_T was also chosen. Outliers were not removed for the same reasons as the global regression model. Tables 4 and 5 give the regression coefficients for each return period for regions 1 and 3, respectively. The general equation is Eq. (5). The same conclusion as for the global model can be drawn regarding residuals.

Regional models give good results (see Table 6 for the performance indices) but do not improve all estimates of AM7\_T; residuals increase for some catchments. The gain



**Table 4.** Regression coefficients of the model for region 1 by return period. The general equation is  $AM7\_T = AREA \times 10^{\text{constant} + ax_1 + bx_2 + \dots + zx_p}$ . Pe is percolation [mm] and SP summer precipitation [mm].

Return period	Constant	Pe	SP
5	1.6924	0.0048	-0.0245
10	1.9390	0.0053	-0.0264
20	2.0912	0.0059	-0.0278
50	2.2464	0.0071	-0.0298

**Table 5.** Regression coefficients of the model for region 3 by return period. The general equation is  $AM7\_T = AREA \times 10^{\text{constant} + ax_1 + bx_2 + \dots + zx_p}$ . ST is summer precipitation [mm], and RC the recession coefficient [ $\text{day}^{-1}$ ].

Return period	Constant	ST	RC
5	-7.7700	0.3482	-12.3893
10	-9.3820	0.4525	-14.9347
20	-11.0117	0.5583	-17.1869
50	-13.6585	0.7312	-20.0026

in precision is not considered sufficient when balanced with the loss of robustness due to the smaller number of catchments used to calibrate and validate the models. The global model is thus preferred at the moment for its greater robustness. However, looking at their current performance, regional models seem promising for the future, when more data from gauging stations will be available. Indeed, it would be interesting to carry out a new cluster analysis and develop one model per homogeneous region in 10 yr. It should actually improve estimates of low flows.

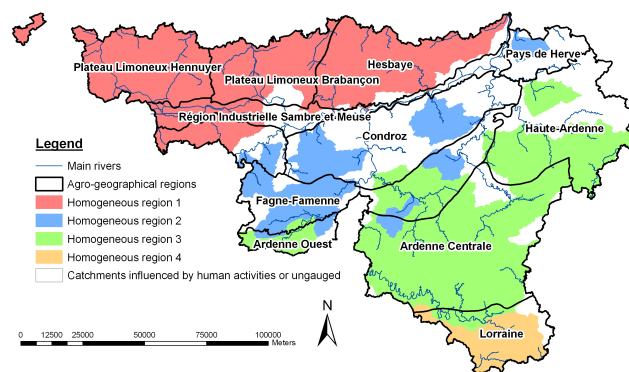
## 4 Discussion

### 4.1 Frequency analysis

Two-parameter lognormal and gamma are the most common distributions that fit low-flow data in Wallonia. The 2-parameter lognormal law has already been used by Galéa et al. (1999) to fit low-flow data of the Loire catchment in France. Gamma was the best distribution for the Missouri catchment in the USA (Joseph, 1970). As our results demonstrated that the type of distribution is not linked to the spatial location or the area of the catchment, a frequency analysis per catchment was essential to determine the best distribution and therefore accurately estimate  $AM7\_T$  for return periods of 5, 10, 20 and 50 yr.

### 4.2 Global regression model

The explanatory variables selected by stepwise for the global model were the recession coefficient and percolation. Indeed,



**Fig. 4.** Map of homogeneous regions and agro-geographical regions in Wallonia. The four homogeneous regions nearly correspond to Walloon agro-geographical regions.

they are the two variables that most correlated to specific  $AM7\_T$  (correlation coefficient of  $-0.47$  and  $0.39$  for a return period of 5 yr) without being correlated to each other (coefficient of  $-0.28$ ). These two features are linked to geology: the more permeable the substratum, the higher the percolation and the lower the recession coefficient. Geology is considered by Smakhtin (2001) as one of the natural factors that most influence low flows. Indeed, the main component of low flow is baseflow, which depends on geology and in particular on substratum permeability. Percolation allows estimating groundwater recharge, and the recession coefficient helps characterise water input from groundwater to the river during low-flow periods. Moreover, Vogel and Kroll (1992) found that low-flow characteristics were highly correlated to catchment area, average basin slope and baseflow recession constant. Therefore, this equation quantifies the role played by geology in determining low flows in Wallonia.

The physical role of the variables selected by the other methods can also be explained. In our region, the precipitation from October to April quantifies water input during the period of groundwater recharge. The hydrological type of soil describes the infiltration rate (high for A to very low for D) and drainage (excellent for A to very bad for D) of soils. Yet, higher infiltration favours higher groundwater recharge, and groundwater is the main source of water in rivers during low-flow periods. Therefore, soils of the hydrological group A permit a higher groundwater input into the rivers during low-flow periods than soils of the hydrological groups B and C. Grasslands favour infiltration thanks to their dense root system. The renewal of the roots creates preferential infiltration paths.

The global model resulted in good prediction for the middle range of specific  $AM7\_T$  ( $4 \times 10^{-4}$  to  $4 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for return periods of 5 and 10 yr,  $3 \times 10^{-4}$  to  $4 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for a return period of 20 yr, and  $2 \times 10^{-4}$  to  $3 \times 10^{-3} \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$  for a return period of 50 yr). Around 10 % of the observations are outside this

**Table 6.** Comparison of regional and global models using performance indices:  $R^2$  (determination coefficient),  $R^2_{\text{adj}}$  (adjusted determination coefficient) and RMSE (root-mean-square error). The numbers in bold are the highest values of indices for each return period.

	Region 1							
	T5		T10		T20		T50	
	Regional model	Global model	Regional model	Global model	Regional model	Global model	Regional model	Global model
Calibration $R^2$	0.868	0.670	0.858	0.623	0.849	0.578	0.842	0.508
Calibration $R^2_{\text{adj}}$	<b>0.850</b>	0.658	<b>0.839</b>	0.610	<b>0.829</b>	0.563	<b>0.821</b>	0.490
Calibration RMSE ( $\text{m}^3 \text{s}^{-1}$ )	<b>0.103</b>	0.201	<b>0.098</b>	0.187	<b>0.094</b>	0.177	<b>0.089</b>	0.170
Validation $R^2$	0.909	0.796	0.884	0.754	0.852	0.703	0.793	0.608
Validation $R^2_{\text{adj}}$	<b>0.849</b>	0.771	<b>0.806</b>	0.724	<b>0.754</b>	0.666	<b>0.656</b>	0.559
Validation RMSE ( $\text{m}^3 \text{s}^{-1}$ )	<b>0.103</b>	0.205	<b>0.105</b>	0.188	<b>0.108</b>	0.179	<b>0.116</b>	0.177

	Region 3							
	T5		T10		T20		T50	
	Regional model	Global model	Regional model	Global model	Regional model	Global model	Regional Rmodel	Global model
Calibration $R^2$	0.888	0.670	0.844	0.623	0.797	0.578	0.731	0.508
Calibration $R^2_{\text{adj}}$	<b>0.875</b>	0.658	<b>0.825</b>	0.610	<b>0.773</b>	0.563	<b>0.699</b>	0.490
Calibration RMSE ( $\text{m}^3 \text{s}^{-1}$ )	<b>0.119</b>	0.201	<b>0.114</b>	0.187	<b>0.110</b>	0.177	<b>0.105</b>	0.170
Validation $R^2$	0.771	0.796	0.759	0.754	0.742	0.703	0.698	0.608
Validation $R^2_{\text{adj}}$	0.706	<b>0.771</b>	0.691	<b>0.724</b>	0.668	0.666	<b>0.611</b>	0.559
Validation RMSE ( $\text{m}^3 \text{s}^{-1}$ )	0.251	<b>0.205</b>	0.208	<b>0.188</b>	0.181	0.179	<b>0.162</b>	0.177

range and have high residuals when predicted by the models. This lack of calibration of the models for very high and low values of specific AM7\_T can explain the low  $R^2$  of the model, and can be solved by adding data, but they are not available yet in Wallonia. It could be useful to recalibrate the models in 10 yr, when more stations will have at least 20 yr of data. However, it is not uncommon to underestimate large flow values by regression models (Laaha and Blöschl, 2006).

The higher performance of the model in validation can be explained by the fact that all validation stations but one have specific AM7 in the middle range of values where the model is best calibrated.

When we plotted the constant and regression coefficients of the models against the return period, we found logarithmic trends. This enabled us to develop a single model for the whole study area, in function of the return period. The use of this equation rather than the equations for each return period does not imply a high loss in precision for AM7\_T estimates (between 0.1 and 3 % for the global model).

This equation can be used by anyone who needs an estimate of low flows at an ungauged site for a desired return period. It will help river managers to improve the management of low flows in rivers, especially regarding water resource planning, reservoir storage design, recreation and environmental flow requirements for wildlife conservation (Veza et al., 2010).

However, one should note that climate change has not been taken into account in this study. Yet, it is forecasted by some studies that low flows will worsen in the future (de Wit et al., 2007; Bauwens et al., 2011). For example, Bauwens et al. (2011) showed that low-flow discharges (AM7) in two Belgian sub-catchments of the Meuse River may decrease by 19–28 % for a return period of 5 yr, and 20–35 % for a 50 yr return period, by the end of the century. The forecasted future worsening of low flows will probably lead to a modification of the equation. Nevertheless, techniques exist to assess the hydrological sensitivity of catchments to climate change (van der Wateren-de-Hoog, 1998, for instance).

#### 4.3 Low-flow regionalisation

We found that Wallonia is composed of four homogeneous regions. Catchments of a same region are contiguous, even though it was not a condition to form the homogeneous regions. It can be thought that X and Y map coordinates have influenced this grouping. However, when these two variables are removed from the cluster analysis, the groups are the same, except for 3 catchments that change group. Nevertheless, the position of the regions follows a north-west-south-east gradient: altitude, slope, precipitation, and percentage of grasslands and forests increase from region 1 to 3, and temperature and percentage of arable lands and soils of good

infiltration rate and drainage decrease. Compared to region 3, region 4 is characterised by lower altitude, precipitation, and percentage of grasslands and forests, and higher slope, temperature, and percentage of arable lands and soils of good infiltration rate and drainage. All these factors are in fact related. Crops grow better on soils of good infiltration rate and drainage, and need suitable temperature and precipitation. Forests rather cover steeper lands situated at higher altitude, where soil properties are less good. Actually, the four regions nearly correspond to natural regions called agro-geographical regions (Fig. 4) (Christians, 1971).

To date, only two regions contain enough gauging stations to permit the development of a specific model. Compared to the global one, they presented little interest because of the loss of robustness. But considering that a lot of new gauging stations have been put in place during the last decade, we propose to reconsider them in the future.

## 5 Conclusions

We developed a full analysis chain allowing us to estimate low flows anywhere in gauged and ungauged catchments in Wallonia, and this for any desired return period between 5 and 50 yr.

This method puts together the selection of gauging stations for low-flow calculation, frequency analysis to fit a frequency distribution to low-flow data, an optional cluster analysis to delineate homogeneous regions if enough data are available, regression analysis to develop models predicting low flows from catchment characteristics, and a new approach that evaluates the relationships between regression coefficients and the return period.

This method answers the need of water managers since it allows them to freely choose the return period of the low flows that they will consider in their regulation for each water use.

This method has been applied to a quite small region. The lack of data available at the moment affected the development of regional models and the precision of the global model for very high and low values of  $AM7_T$ . It is therefore advised to carry out this study again in 10 yr when more stations have at least 20 yr of data. It would also be interesting to take climate change into consideration, and repeat it for other areas to see if the same variables are selected for all return periods.

*Acknowledgements.* This research was funded by the Interreg IVB NEW programme and the Public Service of Wallonia through the AMICE project (Adaptation of the Meuse to the Impacts of Climate Evolutions). Data were made available for free by the Public Service of Wallonia: Operational General Head Office in Mobility and Hydraulic Waterways, Study and Support for Management Department, Head Office of Integrated Hydraulic Management, Hydrology Study Service; and Operational General

Head Office in Agriculture, Natural Resources and Environment, Rurality and Streams Department, Head Office of Non Navigable Streams, AQUALIM (Service public de Wallonie: Direction générale opérationnelle de la Mobilité et des Voies hydrauliques, Département des Études et de l'Appui à la gestion, Direction de la Gestion hydraulique intégrée, Service d'Études Hydrologiques (SETHY); and Direction générale opérationnelle de l'Agriculture, des Ressources naturelles et de l'Environnement, Département de la Ruralité et des Cours d'Eau, Direction des Cours d'eau non navigables, AQUALIM).

Edited by: N. Romano

## References

- Abi-Zeid, I. and Bobée, B.: La modélisation stochastique des étiages: une revue bibliographique, *J. Water Sci.*, 12, 459–484, 1999.
- Bauwens, A., Sohler, C., and Degré, A.: Hydrological response to climate change in the Lesse and the Vesdre catchments: contribution of a physically based model (Wallonia, Belgium), *Hydrol. Earth Syst. Sci.*, 15, 1745–1756, doi:10.5194/hess-15-1745-2011, 2011.
- Chen, Y. D., Huang, G., Shao, Q., and Xu C.-Y.: Regional analysis of low flow using L-moments for Dongjiang Basin, South China, *Hydrolog. Sci. J.*, 51, 1051–1064, 2006.
- Christians, C.: Les degrés de priorité des remembrements de biens ruraux en Belgique: étude géographique des structures et premières propositions de priorités, *Rev. Agr. – Brussels*, 11/12, 1439–1505, 1971.
- Condie, R. and Nix, G. A.: Modelling of low flow frequency distributions and parameter estimation, in: *Proceedings of the International Water Resources Symposium, Water for Arid Lands*, Teheran, Iran, 8–9 December 1975.
- Confais, J. and Le Guen, M.: Premiers pas en régression linéaire avec SAS<sup>®</sup>, *Revue Modulad*, 35, 220–363, available at: <https://www.rocq.inria.fr/axis/modulad/numero-35/Tutoriel-confais-35/confais-35.pdf> (last access: 12 June 2012), 2006.
- Dacharry, M.: Dictionnaire français d'hydrologie, available at: <http://webworld.unesco.org/water/ihp/db/glossary/glu/indexdic.htm>, (last access: 22 February 2012), 1997.
- Dagnélie, P.: *Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions*, Bruxelles, De Boeck, 736 pp., 2006.
- Demarcin, P., Sohler, C., Mokadem, A. I., Dautrebande, S., and Degré, A.: Essai de cartographie des classes d'infiltrabilité des sols de Wallonie (Belgique), *Biotechnol. Agron. Soc.*, 15, 119–128, 2011.
- de Wit, M. J. M., van den Hurk, B., Warmerdam, P. M. M., Torfs, P. J. J. F., Roulin, E., and van Deursen, W. P. A.: Impact of climate change on low-flows in the River Meuse, *Climatic Change*, 82, 351–372, 2007.
- El Adlouni, S., Bobée, B., and Ouarda, T. B. M. J.: On the tails of extreme event distributions in hydrology, *J. Hydrol.*, 355, 16–33, 2008.
- Galéa, G., Mercier, G., and Adler, M. J.: Modèle débit-durée-fréquence d'étiage, concept et usage pour une approche régionale des régimes des basses eaux des bassins hydrographiques de la

- Loire (France) et du Crisu-Alb (Roumanie), *J. Water Sci.*, 12, 93–122, 1999.
- Hayes, D. C.: Low-flow characteristics of streams in Virginia, US Geol. Surv., Water-Supply Paper 2374, 69 pp., 1991.
- Hébert, C., El-Jabi, N., and Caissie, D.: Low flow estimation for New Brunswick Rivers, *Can. Tech. Rep. Fish. Aquat. Sci.*, 2493, 24 pp., 2003.
- Joseph, E. S.: Probability distribution of annual droughts, *J. Irrig. Drain. Div.-ASCE*, 96, 461–474, 1970.
- Kim, S. U. and Lee, K. S.: Regional low flow frequency analysis using bayesian regression and prediction at ungauged catchment in Korea, *KSCE J. Civ. Eng.*, 14, 87–98, 2010.
- Laaha, G. and Blöschl, G.: Low flow estimates from short stream flow records – a comparison of methods, *J. Hydrol.*, 306, 264–286, 2005.
- Laaha, G. and Blöschl, G.: A comparison of low flow regionalisation methods – catchment grouping, *J. Hydrol.*, 323, 193–214, 2006.
- Laaha, G. and Blöschl, G.: A national low flow estimation procedure for Austria, *Hydrolog. Sci. J.*, 52, 625–644, 2007.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and quantiles, *Water Resour. Res.*, 15, 1055–1064, 1979.
- Lang, C. and Gille, E.: Une méthode d'analyse du tarissement des cours d'eau pour la prévision des débits d'étiage, *Norois*, 201, 31–43, available at: <http://norois.revues.org/1743> (last access: 17 December 2012), 2006.
- Leppäjärvi, R.: Frequency analysis of flood and low flow, *IAHS-AISH P.*, 187, 435–442, 1989.
- Mallows, C. L.: More comments on Cp, *Technometrics*, 37, 362–372, 1995.
- Matalas, N. C.: Probability distribution of low flows, *US Geol. Surv., Professional Paper 434-A*, 27 pp., 1963.
- Middelkoop, H., Daamen, K., Gellens, D., Grabs, W., Kwadijk, J. C. J., Lang, H., Parmet, B. W. A. H., Schädler, B., Schulla, J., and Wilke, K.: Impact of climate change on hydrological regimes and water resources management in the Rhine basin, *Climatic Change*, 49, 105–128, 2001.
- Miquel, J.: Guide pratique d'estimation des probabilités de crues, Eyrolles, Paris, France, 160 pp., 1984.
- Modarres, R.: Regional frequency distribution type of low flow in North of Iran by L-moments, *Water Resour. Manag.*, 22, 823–841, 2008.
- Nathan, R. J. and McMahon, T. A.: Practical aspects of low-flow frequency analysis, *Water Resour. Res.*, 26, 2135–2141, 1990.
- Nathan, R. J. and McMahon, T. A.: Estimating low flow characteristics in ungauged catchments, *Water Resour. Manag.*, 6, 85–100, 1992.
- Ouarda, T. B. M. J. and Shu, C.: Regional low-flow frequency analysis using single and ensemble artificial neural networks, *Water Resour. Res.*, 45, 1–16, 2009.
- Pacheco, A., Gottschalk, L., and Krasovskaia, I.: Regionalization of low flow in Costa Rica, in: *Proceedings of the Fifth FRIEND World Conference, Climate Variability and Change – Hydrological Impacts*, Havana, Cuba, November 2006, *IAHS-AISH P.*, 308, 111–116, 2006.
- Saravi, M. M., Abaszadeh, M., Malekian, A., and Samani, A. A. N.: The relationship between morphoclimatic characteristics and peak flows: A case study of the southern Alborz Basins, Iran, *Prog. Phys. Geogr.*, 34, 173–182, 2010.
- Schwarz, G.: Estimating the dimension of a model, *The Annals of Statistics*, 6, 461–464, 1978.
- Smakhtin, V. U.: Low flow hydrology: a review, *J. Hydrol.*, 240, 147–186, 2001.
- Sohier, C., Degré, A., and Dautrebande, S.: From root zone modelling to regional forecasting of nitrate concentration in recharge flows – The case of the Walloon Region (Belgium), *J. Hydrol.*, 369, 350–359, 2009.
- Tallaksen, L. M.: A review of baseflow recession analysis, *J. Hydrol.*, 165, 349–370, 1995.
- Tasker, G. D.: A comparison of methods for estimating low flow characteristics of streams, *Water Resour. Bull.*, 23, 1077–1083, 1987.
- Tsakiris, G., Nalbantis, I., and Cavadias, G.: Regionalization of low flows based on Canonical Correlation Analysis, *Adv. Water Resour.*, 34, 865–872, 2011.
- van der Wateren-de Hoog, B.: A regional model to assess the hydrological sensitivity of medium size catchments to climate variability, *Hydrol. Process.*, 12, 43–56, 1998.
- Veza, P., Comoglio, C., Rosso, M., and Viglione, A.: Low flows regionalization in North-Western Italy, *Water Resour. Manag.*, 24, 4049–4074, 2010.
- Vogel, R. M. and Kroll, C. N.: Low-flow frequency analysis using probability-plot correlation coefficients, *J. Water Res. Pl.-ASCE*, 115, 338–357, 1989.
- Vogel, R. M. and Kroll, C. N.: Regional geohydrologic-geomorphic relationships for the estimation of low-flow statistics, *Water Resour. Res.*, 28, 2451–2458, 1992.
- Ward, J. H.: Hierarchical grouping to optimise an objective function, *J. Am. Stat. Assoc.*, 58, 236–244, 1963.
- Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M., and Wade, A. J.: A review of the potential impacts of climate change on surface water quality, *Hydrolog. Sci. J.*, 54, 101–121, 2009.
- Xanthoulis, D.: Modélisation mathématique des débits d'étiage des cours d'eau: Application au bassin versant de la Lesse, Ph.D. thesis, Université de Liège – Gembloux Agro-Bio Tech, Belgium, 188 pp., 1985.
- Yue, S. and Pilon, P.: Probability distribution type of Canadian annual minimum streamflow, *Hydrolog. Sci. J.*, 50, 427–438, 2005.