Hydrol. Earth Syst. Sci. Discuss., 9, 11583–11614, 2012
www.hydrol-earth-syst-sci-discuss.net/9/11583/2012/
doi:10.5194/hessd-9-11583-2012

Hydrology and Earth System Sciences Discussions

A method for low flow estimation at ungauged sites, 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

Received: 2 October 2012 – Accepted: 2 October 2012 – Published: 11 October 2012

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

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Being able to estimate low flows at any point of a river is really important nowadays for a good integrated management of rivers. Knowing the magnitude as well as the frequency of such extreme events becomes essential. In order to build a model of low flow calculation, usable in ungauged catchments and which takes also into account low flow frequency, we started with a low flow frequency analysis including a comparison of different distributions. Two-parameter Lognormal and Gamma were the most common distributions that fit low flow data in Wallonia. This was followed by a regionalisation of low flows using 25 different climatic and physical catchment variables, and the development of regression models that can be used to estimate the minimum 7-day average flow for different return periods, using catchment characteristics. The variables the most correlated to specific minimum 7-day average flows were the recession coefficient and percolation, 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. Finally, regression coefficients were logarithmically linked to the return period. This enabled us to develop a single model per region and for the whole study area, in function of the return period. In conclusion, the method developed in this study allows us to estimate low flows in gauged and ungauged catchments of a given region for a given return period. The interest of regionalisation and development of regional models is also discussed.

1 Introduction

Low flows can have different meanings depending 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.

HESSD
9, 11583–11614, 2012

The knowledge of river behaviour during low flows is really important nowadays for a good integrated management of rivers (managing water quality in regards to the Water Framework Directive, biodiversity, water resources, etc.). Low flow estimates are needed in projects relating, for example, to the construction of hydroelectric dams, the conservation of aquatic habitats, water abstraction licensing, discharge management, commercial navigation or the management of bathing areas (Smakhtin, 2001; WMO, 2008). Furthermore, it becomes essential to know the frequency of such extreme events in addition to their magnitude. The frequency actually indicates the severity of low flow events. And, in case of droughts, different measures to maintain a minimum flow in rivers are usually taken by the environmental department of the Government according to this severity (e.g. "Plan d'Action Sécheresse" in France, drought plans in the UK).

Low flow frequency analysis aims at improving the accuracy of low flow estimation for any return period (T). Indeed, the amount of observed data is usually not enough to properly quantify the frequency of extreme events. Thus, a distribution needs to 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, 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).

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. This method 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. This has been done by numbers of scientists such as Nathan and McMahon (1992) for

HESSD
9, 11583–11614, 2012

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 but 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.

The only ones to link frequency analysis and estimation of low flows for ungauged sites were Chen et al. (2006), Kim and Lee (2009), Ouarda and Shu (2009), and Hayes (1991). However, Chen et al. (2006) did not carry out any delineation of homogeneous regions and they only used catchment area as a characteristic to describe low flows for regression. Kim and Lee (2009) used Bayesian multiple regression analysis for regional low flow frequency and utilised the developed model to predict low flow at ungauged sites but only for a 10-yr return period. After a low flow frequency analysis, Ouarda and Shu (2009) developed models to estimate low flows in Quebec but only for three different return periods and taking into account seven physiographical and meteorological variables. Hayes (1991) did not compare distributions and used Pearson type III for frequency analysis. Then, he developed regional models to estimate the annual minimum average 7-consecutive-day discharge for only two different return periods at ungauged sites. 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. Nevertheless, they only considered seven climatic and physical catchment characteristics.

To sum up, none of them developed a model that can be used to estimate low flows in any ungauged river for any desired return period. The aim of this study is therefore to fill this gap. Consequently, we propose a full analysis chain: a low flow frequency analysis with comparison of different distributions is first carried out, followed by a regionalisation of low flows using 25 different variables, and the development of a regression model that can be used to estimate the minimum 7-day average flow for any return period.

2 Material and methods

2.1 Study area

The Walloon Region of Belgium covers an area of 16 844 km^2. 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^2).

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 index, was selected for this study. The main advantages of this parameter are that it eliminates day-to-day variations and allow analyses to be less sensitive to measurement errors. Moreover, 7-day low flows are not very different from 1-day low flows (Smakhtin, 2001). Doing the average over some days also allows avoiding some human influences on flows.

2.3 Selection of gauging stations

More than 240 gauging stations have been installed in Wallonia during the last 40 yr. Out of these stations, we selected those which fulfilled several criteria:

– minimum 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,

HESSD
9, 11583–11614, 2012

– MAM7 over 5 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.

Fifty-nine gauging stations were eventually selected. The main rejection reason was a too short chronicle.

2.4 Frequency analysis

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 (7Q2 and 7Q10 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 45 yr), AM7_T cannot be predicted with accuracy for 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 gave 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 which were approximated via Schwarz method. As no a priori information on the suitability of each law was available, prior probabilities were considered equals 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 χ^2 test was also carried out to check the adjustment of each distribution to the sample.

Finally, Fréchet distribution and the three best distributions for which the χ^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).

This selection of the best distribution, except for 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).

2.5 Low flow regionalisation

It is admitted that catchments which have similar physical and climatic features have similar hydrological responses (Smakhtin, 2001). In this study, climatic and physical catchment characteristics were described by 25 variables: altitude of the gauging station (Alt [m]), map coordinates of the station (X and Y [m]), catchment area (A [km^2]), drainage density (DD [$km\,km^{-2}$]), 10th, 50th and 90th percentiles of slope (Sl_{10}, Sl_{50} and Sl_{90} [%]), area percentages of urban lands (L_u [%]), forests (L_f [%]), arable lands (L_a [%]), permanent crops (L_p [%]) and grasslands (L_g [%]), area percentages of soils of hydrological group A (S_A [%]), B (S_B [%]), C (S_C [%]), and D (S_D [%]) and soils that were not mapped (S_{NM} [%]), annual, summer and winter precipitation (AP, SP and WP [mm]), summer temperature (ST [°C]), potential evapotranspiration (PET [mm]), percolation (Pe [mm]) and recession coefficient (RC [day^{-1}]). “Summer” refers to the July to September period and “winter” refers to the October to April period. 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. 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, Sls, Ls and Ss) were derived from GIS maps. Meteorological data (AP, SP, WP, ST), PET and Pe were simulated by the hydrological model EPICgrid (Sohier et al., 2009). Recession coefficients were calculated using the method developed by Lang and Gille (2006) but adapted to Wallonia.

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 as the merging strategy. Moreover, this method was often 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 of R^2 between two groupings. Indeed, R^2 , the determination coefficient, represents the proportion of information kept after the fusion of clusters. The fusions were stopped before this difference of R^2 became large.

Then, 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 compare groups according to these 25 variables.

HESSD
9, 11583–11614, 2012

2.6 Development of regression models

Since catchment area is highly correlated to AM7_T (correlation coefficient of 0.72 for $T = 5$ yr), we used specific flows (AM7_T 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 of the variability of specific AM7_T. 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 Mallow coefficient (C_p) close to $p+1$ was selected. 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_T 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 dataset associated to 19 stations not selected earlier because of their too short length of data.

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):

HESSD
9, 11583–11614, 2012

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

$$R^2_{\mathrm{adj}} = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right), \tag{2}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2} \tag{3}$$

in which var(Y) is the variance of observed AM7_T, n is the number of observations, p is the number of variables in the model, Y_i is the observed value of AM7_T for the observation i and $\hat{Y}_i$ the predicted value.

All statistical analyses were performed using SAS (Statistical Analysis System) software.

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

3 Results

3.1 Frequency analysis

For each of the 59 stations, the distribution that best fits 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 either the length of data, catchment area or the spatial location of the catchment.

HESSD
9, 11583–11614, 2012

3.2 Low flow regionalisation

The cluster analysis gave four groups of catchments. Catchments in a homogeneous region are contiguous.

The PCA helped understand the differences between groups. As it can be seen in Fig. 1, 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, gentle slopes, receiving less precipitation, where summer temperatures are higher, rather agricultural and urban, with soils of good infiltration capacity and permeability predominating.

Region 2 has a central spatial location. It is a region characterised mainly by intermediate values of features: medium altitude, steep slopes, receiving average precipitation, where summer temperatures are average for Belgium, fairly urbanised and agricultural, rather forested and grassy, with soils of moderate infiltration capacity and relatively low permeability predominating.

Region 3 is situated in the south of Wallonia. It is a region of higher altitude, steep slopes, receiving a lot of precipitation, where summer temperatures are lower, not much urbanised, with soils of relatively good infiltration capacity and moderate permeability predominating, where forests and grasslands prevail.

Region 4 is in the south of region 3. It is a region of higher altitude, steep slopes, receiving a lot of precipitation, where summer temperatures are average, not much urbanised, rather forested and grassy, with soils of low infiltration capacity but good permeability predominating.

3.3 Development of regression models

3.3.1 Global model

First, a global model for the whole study area was developed.

HESSD
9, 11583–11614, 2012

Applicability conditions were checked and the logarithm of AM7_T 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_T.

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 gave the same final equations, except for the 50-yr return period:

$$\mathrm{AM7\_T5} = \mathrm{Area}10^{-3.8935+0.0080L_\mathrm{g}+0.0098S_\mathrm{A}+0.0017\mathrm{WP}-19.8274\mathrm{RC}} \tag{4}$$

$$\mathrm{AM7\_T10} = \mathrm{Area}10^{-3.9741+0.0089L_\mathrm{g}+0.0104S_\mathrm{A}+0.0017\mathrm{WP}-21.2826\mathrm{RC}} \tag{5}$$

$$\mathrm{AM7\_T20} = \mathrm{Area}10^{-4.0639+0.0098L_\mathrm{g}+0.0110S_\mathrm{A}+0.0017\mathrm{WP}-22.4033\mathrm{RC}} \tag{6}$$

$$\mathrm{AM7\_T50} = \mathrm{Area}10^{-4.4764+0.0348Sl_{90}+0.0154S_\mathrm{A}+0.0016\mathrm{WP}-20.4359\mathrm{RC}} \tag{7}$$

for maximum R^2 improvement

$$\mathrm{AM7\_T50} = \mathrm{Area}10^{-3.1899+0.0409Sl_{90}-0.0148S_\mathrm{B}-0.0183S_\mathrm{C}+0.0017\mathrm{WP}-19.1631\mathrm{RC}} \tag{8}$$

for adjusted R^2 selection.

The stepwise method gave the following final equations:

$$\mathrm{AM7\_T5} = \mathrm{Area}10^{-2.7851+0.0017Pe-13.4274RC} \tag{9}$$

$$\mathrm{AM7\_T10} = \mathrm{Area}10^{-2.8396+0.0016\mathrm{Pe}-15.5904\mathrm{RC}} \tag{10}$$

$$\mathrm{AM7\_T20} = \mathrm{Area}10^{-2.8939+0.0016\mathrm{Pe}-17.3211\mathrm{RC}} \tag{11}$$

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

$$\text{AM7\_T50} = \text{Area}10^{-2.9808+0.0016\text{Pe}-19.0192\text{RC}} \tag{12}$$

with L_g the area percentage of grasslands [%]; S_A the area percentage of soils of hydrological group A [%]; WP winter precipitation [mm]; RC recession coefficient [day^{-1}]; SL_{90} the 90th percentile of slope [%]; S_B the area percentage of soils of hydrological group B [%]; S_C the area percentage of soils of hydrological group C [%]; Pe percolation [mm].

Table 1 gives, for each method, the values of R^2, adjusted R^2 and RMSE for the calibration and the validation of the models. R^2 gives the part of the variability of AM7_T explained by the model. Adjusted R^2 takes into account the number of variables and is useful to compare different models. RMSE quantifies the difference between observed and predicted AM7_T.

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 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.

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 it can be observed on Fig. 2 for T5. Therefore, the models overestimate low values of specific AM7_T (especially under $10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5 and T10, and under $5 \times 10^{-4}\,m^3\,s^{-1}\,km^{-2}$ for T20 and T50), and underestimate higher values of specific AM7_T (especially over $3 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5 and T10, and over $2 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T20 and T50). This problem of lack of calibration for extremes is due to the small number of observations for this range of specific AM7_T, especially for specific AM7_T over $4 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5, T10 and T20 and over $3 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T50.

Plotting the constant and regression coefficients of the models against the return period, Fig. 3 shows that the constant and regression coefficients are linked to the return period by a logarithmic relationship.

It is therefore possible to calculate AM7_T for any return period T with this formula:

$$\mathrm{AM7\_T} = \mathrm{Area}10^{-2.6457-0.0847\ln T+0.0017\mathrm{Pe}-9.8077\mathrm{RC}-4\times10^{-5}\mathrm{Pe}\ln T-2.4148\mathrm{RC}\ln T} \tag{13}$$

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, 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. This is also sometimes the case for the 20-yr return period.

HESSD
9, 11583–11614, 2012

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

3.3.2 Regional models

It was shown by several studies that developing one regression equation per region gives better estimates than a global equation (Smakhtin, 2001; Laaha and Blöschl, 2006; Vezza 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.

Nevertheless, we developed regional models by stepwise 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.

The stepwise method gave the following final equations:

For region 1:

$$\mathrm{AM7\_T5} = \mathrm{Area}10^{1.6924+0.0048\mathrm{Pe}-0.0245\mathrm{SP}} \tag{14}$$

$$\mathrm{AM7\_T10} = \mathrm{Area}10^{1.9390+0.0053\mathrm{Pe}-0.0264\mathrm{SP}} \tag{15}$$

$$\mathrm{AM7\_T20} = \mathrm{Area}10^{2.0912+0.0059\mathrm{Pe}-0.0278\mathrm{SP}} \tag{16}$$

$$\mathrm{AM7\_T50} = \mathrm{Area}10^{2.2464+0.0071\mathrm{Pe}-0.0298\mathrm{SP}} \tag{17}$$

For region 3:

$$\mathrm{AM7\_T5} = \mathrm{Area}10^{-7.7700+0.3482\mathrm{ST}-12.3893\mathrm{RC}} \tag{18}$$

$$\mathrm{AM7\_T10} = \mathrm{Area}10^{-9.3820+0.4525\mathrm{ST}-14.9347\mathrm{RC}} \tag{19}$$

$$\mathrm{AM7\_T20} = \mathrm{Area}10^{-11.0117+0.5583\mathrm{ST}-17.1869\mathrm{RC}} \tag{20}$$

HESSD
9, 11583–11614, 2012

$$AM7\_T50 = Area10^{-13.6585+0.7312ST-20.0026RC} \quad (21)$$

with Pe percolation [mm]; SP summer precipitation [mm]; ST summer temperature [°C]; RC recession coefficient [day^{-1}].

As it can be seen on Fig. 4, residuals of regional models increase with observed specific AM7_T but the linear relationships are less marked than the one for the global model.

With regards to region 3, no flow exceeds $3 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5 and T10, and $2 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T20 and T50. However, below an observed specific AM7_T of $10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5 and T10, and $5 \times 10^{-4}\,m^3\,s^{-1}\,km^{-2}$ for T20 and T50, nearly all residuals are negative (overestimation of AM7_T by the models). Over an observed specific AM7_T of $1.5 \times 10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T5 and T10, and $10^{-3}\,m^3\,s^{-1}\,km^{-2}$ for T20 and T50, all residuals are positive (underestimation).

Likewise the global model, the constant and regression coefficients are logarithmically linked to the return period, as shown by Fig. 5.

We therefore obtained these equations:

For region 1:

$$AM7\_T = Area10^{1.3537+0.2361\ln T+0.0031Pe-0.0209SP+0.001Pe\ln T-0.0023SP\ln T} \quad (22)$$

For region 3:

$$AM7\_T = Area10^{-3.5601-2.5492\ln T+0.0741ST-7.2218RC+0.1658ST\ln T-3.2927RC\ln T} \quad (23)$$

For region 1, estimated values of AM7_T obtained by this model are generally higher, in comparison to the estimates obtained by the model for each return period. AM7_T that are already overestimated by the models for each return period are therefore slightly overestimated, and underestimated AM7_T estimates are improved. The difference between estimates ranges from 0.2 to 2.5 % for a return period of 5 yr but increases with the return period to 2 to 4 % for a 50-yr return period.

For region 3, for return periods of 5 and 50 yr, half of estimates are higher and half are lower. The difference is between 0.5 and 2.5 %. For 10-yr and 20-yr return periods, estimates are generally higher by 0.1 to 2.5 %.

HESSD
9, 11583–11614, 2012

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 low 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 Low flow regionalisation

We found that 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. But, 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, percentage of grasslands and forests increase from region 1 to 3, and temperature, 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, percentage of grasslands and forests, and higher slope, temperature, 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 in higher altitude, where soil properties are less good. Actually, the four regions nearly correspond to natural regions called agro-geographical regions (Fig. 6): region 1 to the loamy low plateaux ("Plateau limoneux hennuyer", "Plateau

HESSD
9, 11583–11614, 2012

limoneux brabançon" and "Hesbaye"), region 2 to "Condroz", "Fagne-Famenne" and "Pays de Herve", region 3 to "Ardenne" and region 4 to "Lorraine belge". These agro-geographical regions were delimitated according to geology, types of soil (fertility) and relief (Christians, 1971). The geology influences the type of soils and they both have an influence on land use. The climate of a region is partly determined by its relief. As our study accounted for these factors through more specific parameters, this correspondence is not surprising.

4.3 Models

The explanatory variables selected by stepwise for the global model were the recession coefficient and percolation. Indeed, they are the two variables the most correlated to specific AM7_T (correlation coefficient of −0.47 and 0.39, respectively, 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 is, the higher percolation is and the lower the recession coefficient is. And geology is considered by Smakhtin (2001) as one of the natural factors that most influence low flows. 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.

Regarding regional models, percolation and summer precipitation were the variables selected for region 1 (correlation coefficient with AM7s_T5 of 0.64 and −0.34, respectively). The recession coefficient was actually more correlated to AM7s_T5 than summer precipitation (coefficient of −0.44) but was not selected because it was also correlated to percolation (coefficient of −0.67). Summer temperature and recession coefficient were selected for region 3 (correlation coefficient with AM7s_T5 of −0.44 and 0.37, respectively). Percolation was not highly correlated to AM7s_T5 for this group (coefficient of −0.01).

The global model and the model for region 1 give good prediction for the middle range of specific AM7_T (4×10^{-4} to 4×10^{-3} m^3 s^{-1} km^{-2} for return periods of 5 and

HESSD
9, 11583–11614, 2012

10 yr, 3×10^{-4} to 4×10^{-3} $m^3 s^{-1} km^{-2}$ for a return period of 20 yr, and 2×10^{-4} to 3×10^{-3} $m^3 s^{-1} km^{-2}$ for a return period of 50 yr). Around 10 % of the observations are outside this range and have high residuals when predicted by the models. This lack of calibration of the models for extremes 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). For region 3, the range of values of specific AM7_T in which the model is more precise is not as clear but no flow exceeds 3×10^{-3} $m^3 s^{-1} km^{-2}$ for T5 and T10, and 2×10^{-3} $m^3 s^{-1} km^{-2}$ for T20 and T50. A recalibration of this model with more data can also be done since it will improve its predicting performance anyway.

When we plotted the constant and recession coefficients of the models against the return period, we found logarithmic trends. This enabled us to develop a single model for region 1, region 3, and 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, and between 0.1 and 4 % for regional models).

4.4 Comparison of regional and global models

Table 2 shows the different performance indices (R^2, R^2_{adj} and RMSE) for the calibration and the validation of regional and global models.

For region 1, the regional model is always more performing than the global model. The part of the variance explained by the regional model in calibration and validation as well as the residuals of calibration decrease when the return period increases. The residuals of validation increase slightly with the return period.

For region 3, the regional model has a greater performance in calibration, but not in validation except for a 50-yr return period. When the return period increases, the validation and calibration R^2 and RMSE of the regional model diminish.

HESSD
9, 11583–11614, 2012

When comparing the performance and behaviour of the regional models, no clear general trend for the two regions can be highlighted. Regional models are actually sensitive to the particularities of the catchments included in the region, which is mainly due to the small number of gauging stations (for calibration and validation, respectively, 18 and 6 for region 1, 20 and 10 for region 3). The global model, calibrated and validated with more stations (59 and 19, respectively), reduces the noise created by some stations and is therefore more robust.

Regional models give good results but do not improve all estimates of AM7_T, residuals increase for some catchments. The gain 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 then 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.

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 then help river managers to improve the management of low flows in rivers, especially regarding water resources planning, reservoir storage design, recreation and environmental flow requirements for wildlife conservation (Vezza 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 to 28 % for a return period of 5 yr, and 20 to 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

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

assess the hydrological sensitivity of catchments to climate change (van der Wateren-de-Hoog, 1998 for instance).

5 Conclusions

We developed a full analysis chain allowing us to estimate low flows anywhere in gauged and ungauged catchments of a given region and this, for any desired return period. Knowing the magnitude and the frequency of such extreme events will help managers to improve the management of low flows in rivers and the management of droughts.

Indeed, this method is very complete. It puts together a method of selection of gauging stations for low flow calculation, frequency analysis to fit a frequency distribution to low flow data, cluster analysis to delineate homogeneous regions, and regression analysis to develop models predicting low flows from catchment characteristics for any return period. To our knowledge, this kind of study that joins frequency analysis and regionalisation has never been done in such a complete way for low flows.

This method has been applied to a small region. The lack of data available at the moment affected the development of regional models and the precision of the global model for extremes. 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 program 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,

HESSD
9, 11583–11614, 2012

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'Etudes 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).

References

Abi-Zeid, I. and Bobée, B.: La modélisation stochastique des étiages: une revue bibliographique, J. Water Sci., 12, 459–484, 1999.

Bauwens, A., Sohier, 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, Hydrol. 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®, Revue Modulad, 35, 220–363, available at: https://www.rocq.inria.fr/axis/modulad/numero-35/Tutoriel-confais-35/confais-35.pdf, 2006.

Dacharry, M.: Dictionnaire français d'hydrologie: http://webworld.unesco.org/water/ihp/db/glossary/glu/indexdic.htm (last access: 22 February 2012), 1997.

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.

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.
Joseph, E. S.: Probability distribution of annual droughts, J. Irr. Drain. Div.-ASCE, 96 (IR4), 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, 2006.
Leppäjärvi, R.: Frequency analysis of flood and low flow, IAHS-AISH P., 187, 435–442, 1989.
Matalas, N. C.: Probability distribution of low flows, US Geol. Surv., Professional Paper 434-A, 27 pp., 1963.
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, 2135–2041, 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.
Saravi, M. M., Abasizadeh, 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. Geog., 34, 173–182, 2010.

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

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.
Vezza, 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-geomorpic relationships for the estimation of low-flow statistics, Water Resour. Res., 28, 2451–2458, 1992.
World Meteorological Organisation: Manual on low-flow: Estimation and prediction, Operational Hydrology Report 50, 136 pp., 2008.
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.

HESSD
9, 11583–11614, 2012

Table 1. Performance of the models according to different indices: R^2 (determination coefficient), R^2_{adj} (adjusted determination coefficient) and RMSE (Root-mean-square error). The stepwise model seems the best option for the purpose of this study.

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

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

Table 2. Comparison of regional and global models using performance indices: R^2 (determination coefficient), R^2_{adj} (adjusted determination coefficient) and RMSE (Root-mean-square error).

	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_{adj}	**0.850**	0.658	**0.839**	0.610	**0.829**	0.563	**0.821**	0.490
Calibration RMSE ($m^3 s^{-1}$)	**0.103**	0.201	**0.098**	0.187	**0.094**	0.177	**0.089**	0.170
Validation R^2	0.909	0.796	0.884	0.754	0.852	0.703	0.793	0.608
Validation R^2_{adj}	**0.849**	0.771	**0.806**	0.724	**0.754**	0.666	**0.656**	0.559
Validation RMSE ($m^3 s^{-1}$)	**0.103**	0.205	**0.105**	0.188	**0.108**	0.179	**0.116**	0.177
	Region 3							
	T5		T10		T20		T50	
	Regional model	Global model	Regional model	Global model	Regional model	Global model	Regional model	Global model
Calibration R^2	0.888	0.670	0.844	0.623	0.797	0.578	0.731	0.508
Calibration R^2_{adj}	**0.875**	0.658	**0.825**	0.610	**0.773**	0.563	**0.699**	0.490
Calibration RMSE ($m^3 s^{-1}$)	**0.119**	0.201	**0.114**	0.187	**0.110**	0.177	**0.105**	0.170
Validation R^2	0.771	0.796	0.759	0.754	0.742	0.703	0.698	0.608
Validation R^2_{adj}	0.706	**0.771**	0.691	**0.724**	0.668	0.666	**0.611**	0.559
Validation RMSE ($m^3 s^{-1}$)	0.251	**0.205**	0.208	**0.188**	0.181	0.179	**0.162**	0.177

HESSD
9, 11583–11614, 2012

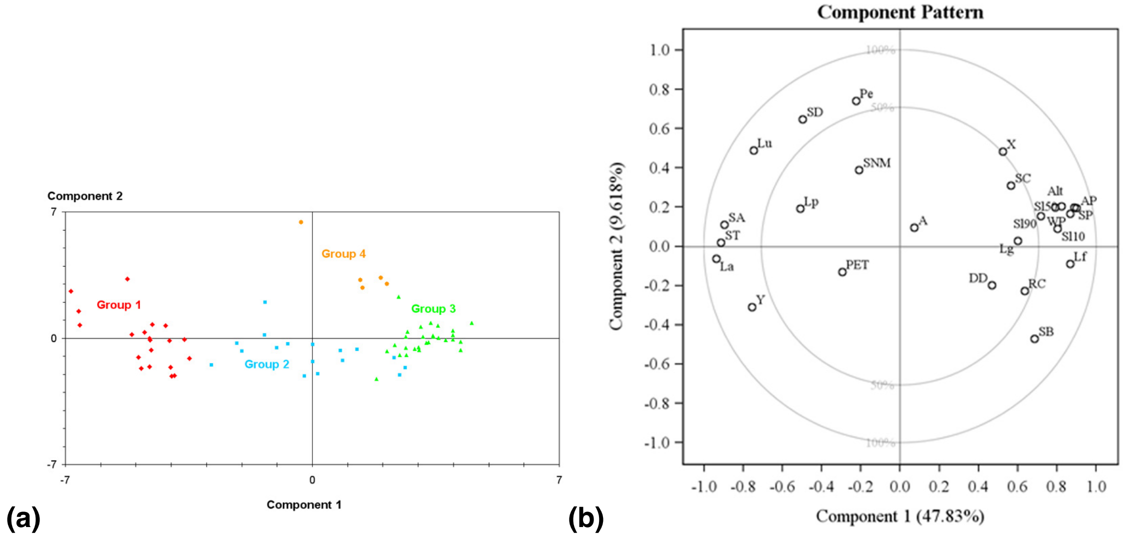

Fig. 1. Groups of catchments **(a)** and correlation circle **(b)** in the plane of the two first principal components. The two first components of the PCA allow distinguishing the four groups of catchments.

HESSD
9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD

9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

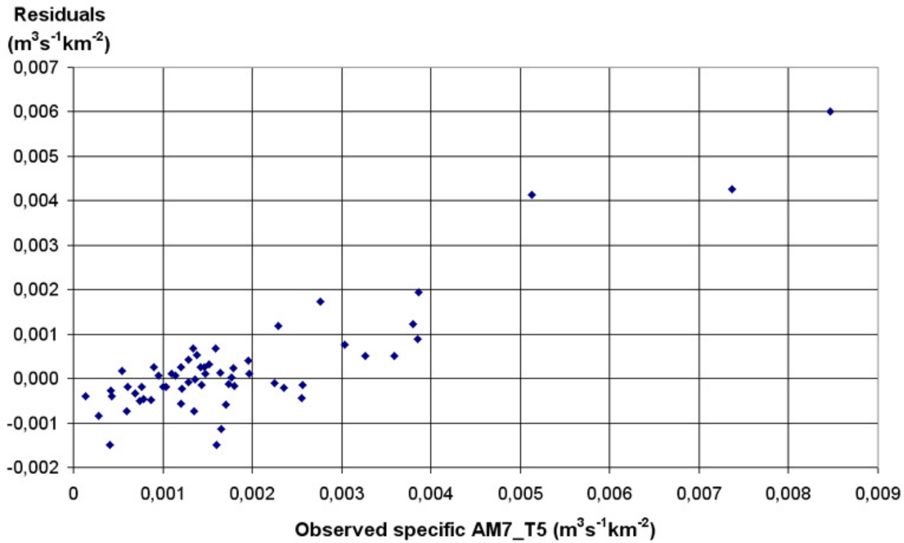

Fig. 2. Plot of global model residuals against observed specific AM7_T (example for the return period of 5 yr). Residuals increase linearly with observed specific AM7_T.

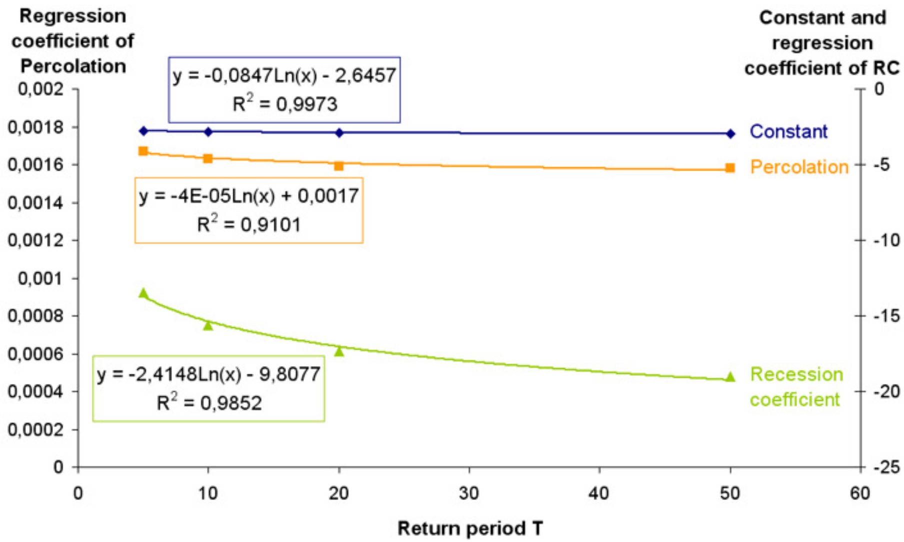

Fig. 3. Constant and regression coefficients of the global model plotted against the return period. There is a logarithmic relationship between the constant and regression coefficients, and the return period.

HESSD
9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

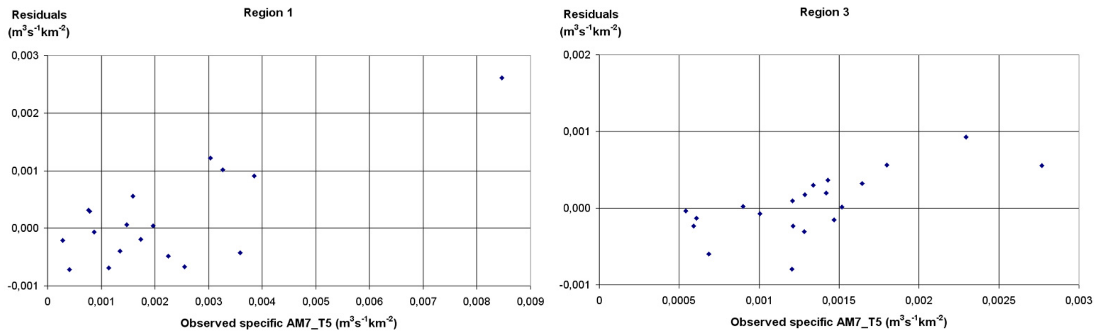

Fig. 4. Plots of regional model residuals against observed specific AM7_T (example for the return period of 5 yr). Residuals increase linearly with observed specific AM7_T.

HESSD
9, 11583–11614, 2012

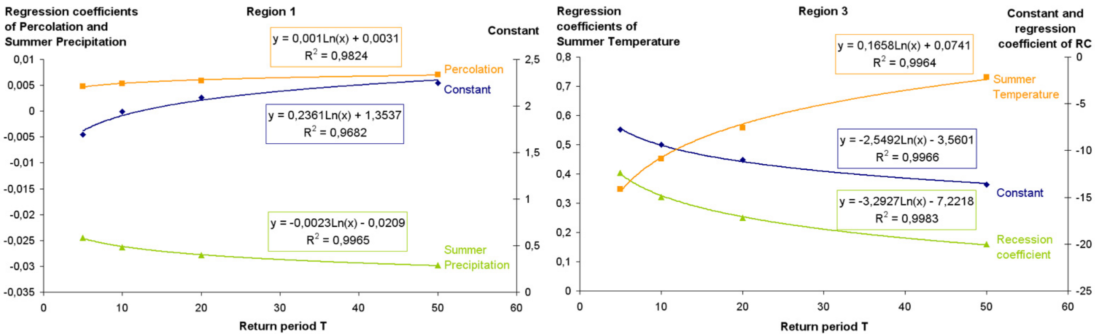

Fig. 5. Constant and regression coefficients of the regional models plotted against the return period. There is a logarithmic relationship between the constant and regression coefficients, and the return period.

HESSD
9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.

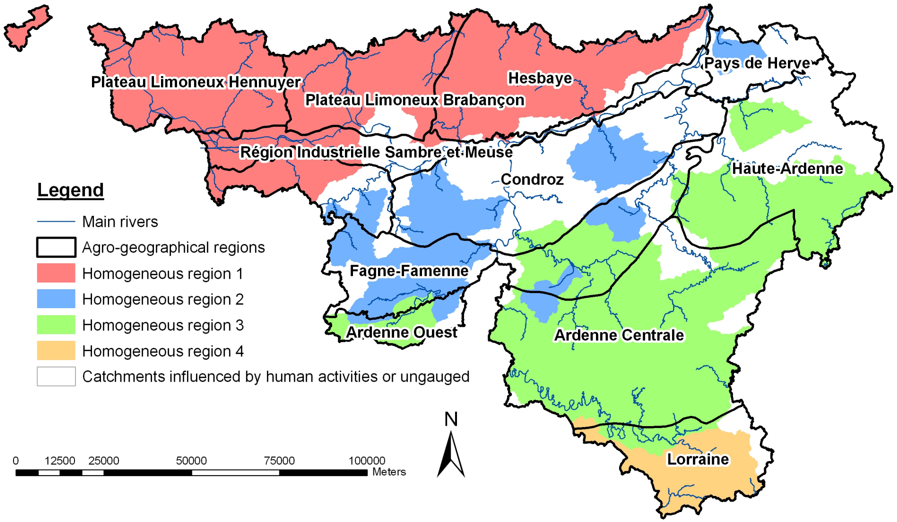

Fig. 6. Map of homogeneous regions and agro-geographical regions in Wallonia. The four homogeneous regions nearly correspond to Walloon agro-geographical regions.

HESSD
9, 11583–11614, 2012

A method for low flow estimation at ungauged sites

M. Grandry et al.