Assessing winter storm flow generation by means of permeability of the lithology and dominating runoff production processes

H. Hellebrand, L. Hoffmann, J. Juilleret, L. Pfister

**Hydrology and
Earth System
Sciences**

Assessing winter storm flow generation by means of permeability of the lithology and dominating runoff production processes

H. Hellebrand, L. Hoffmann, J. Juilleret, and L. Pfister

Public Research Center-Gabriel Lippmann, Belvaux, Grand Duchy of Luxembourg

Abstract. In this study two approaches are used to predict winter storm flow coefficients in meso-scale basins ($10 \, \text{km}^2$ to $1000 \, \text{km}^2$) with a view to regionalization. The winter storm flow coefficient corresponds to the ratio between direct discharge and rainfall. It is basin specific and supposed to give an integrated response to rainfall. The two approaches, which used the permeability of the substratum and dominating runoff generation processes as basin attributes are compared. The study area is the Rhineland Palatinate and the Grand Duchy of Luxembourg and the study focuses on the Nahe basin and its 16 sub-basins (Rhineland Palatinate). For the comparison, three statistical models were derived by means of regression analysis. The models used the winter storm flow coefficient as the dependent variable; the independent variables were the permeability of the substratum, preliminary derived dominating runoff generation processes and a combination of both. It is demonstrated that the permeability and the preliminary derived processes carry different layers of information. Cross-validation and statistical tests were used to determine and evaluate model differences. The cross-validation resulted in a best model performance for the model that used both parameters, followed by the model that used the dominant runoff generation processes. From the statistical tests it was concluded that the models come from different populations, carrying different information layers. Analysis of the residuals of the models indicated that the permeability and runoff generation processes did provide complementary information. Simple linear models appeared to perform well in describing the winter storm flow coefficient at the meso-scale when a combination of the permeability of the substratum and dominating runoff generation processes served as independent parameters.

1 Introduction

Regionalization is a widely used procedure in hydrology (Burn, 1997; Post and Jakeman, 1999; Kokkonen et al., 2003; Croke and Norton, 2004; Merz and Blöschl, 2004; Parajka et al., 2005; Merz et al., 2006), regionalization being defined as the transfer of information from one basin to another (Blöschl and Sivapalan, 1995). Regression analysis is the most widely used regionalization technique, although alternative techniques are also used (Kokkonen et al., 2003). Since a regression needs a dependent and at least one independent variable, the choice of the variables is usually a hydrological variable as dependent and one or several physiographic basin characteristics as independent variables. Mazvimavi (2003) listed the most commonly used physiographic basin characteristics in regression analyses, which are: land use, geology, drainage density and basin area. Pfister et al. (2002) developed a methodology that determines the qualitative behavior of gauged basins with short historical data series with a view to regionalization, using the winter storm flow coefficient, or C-value, as the dependent parameter in a regression analysis. The previously named basin characteristics served as independent parameters. The C-value is defined as the ratio between storm flow and rainfall, is supposed to be basin specific and to have a strong seasonal variability, and should be more or less constant during winter, expressing the saturated state of the basin (Pfister et al., 2002). Uhlenbrook et al. (2004) pointed out that in meso-scale basins processes combine into a more complex way, producing an integrated runoff response to rainfall. In this study, the C-value is supposed to represent this response of meso-scale basins (i.e. basins ranging in size from $10 \, \text{km}^2$ to $10^3 \, \text{km}^2$; Blöschl, 1996) to rainfall during winter. The permeability of the substratum was found to be an important basin characteristic in describing the C-value of basins in the Grand Duchy of Luxembourg (Pfister et al., 2002) and it will serve in this study as a single independent parameter

Correspondence to: H. Hellebrand
(hellebra@lippmann.lu)

in a first model, which is based on linear regression, to describe the C-value with. Since soils form the first medium between precipitation and runoff generating processes after the vegetation cover, they will serve in this study as independent parameters for a second model.

The impermeability of the substratum, which is used in this study, is based on a methodology of Zumstein et al. (1989), who classified the infiltration permeability of the substratum with respect to its lithology and geohydrological characteristics such as fractures and porosity, obtaining eight different permeability classes. This classification was adapted and simplified into only three classes: permeable, semi-permeable and impermeable. The determination of the three classes in this study was based on the available digitized geological map (Geologische Übersichtskarte, scale 1:300 000 (GÜK 300)). Due to this simplification and due to the coarse scale of the geological map, the thus assessed permeability should be regarded as an indicator for runoff production only.

Scherrer and Naef (2003) developed an approach to determine runoff processes at the plot scale. It uses soil data, geology, topography and vegetation for the process identification. Scherrer (1997) and Faeh (1997) conducted sprinkling experiments in Switzerland on grassland hill slopes with varying slopes, geology and soils and recorded the soil-water levels, soil-water content and soil-water tension. The outcome of this research formed the basis for developing process decision schemes, which reflect the complex nature of runoff formation to eventually determine the dominating runoff generation process on a soil profile (Scherrer and Naef, 2003). Since several runoff processes can occur at the same site, the one process that contributes most is called the dominating runoff generation process. Schmocker-Fackel et al. (2007) used this approach for identifying runoff processes at the plot as well as at catchment scale and illustrated the potential of the use of dominating runoff generation processes for defining the infiltration parameters used in rainfall-runoff models. However, the approach of Scherrer and Naef (2003) is time consuming and often, detailed soil data is lacking to apply it on a smaller scale. The approach has been up-scaled using an artificial neural network (ANN) model developed by Steinrücken et al. (2006). This model was applied to the Nahe basin, resulting in a digitized map that provides the preliminary modelled dominating runoff generation processes for this basin. The results of Steinrücken et al. (2006) formed the basis for the derivation of the processes that are used in this study.

The dominating runoff generation processes obtained from Steinrücken et al. (2006) are: Saturated Overland Flow (D_{SOF}), SubSurface Flow (D_{SSF}) and Deep Percolation (D_{DP}). The SOF and SSF processes are subdivided into D_{SOF1}, D_{SOF2} and D_{SOF3} and D_{SSF1}, D_{SSF2} and D_{SSF3}. The numbers refer to the intensity of which the processes react to rainfall, where 1 has relatively the most abruptly changing flow reaction and 3 the most gradually changing flow re-

action. The "detailedness" of these processes is supposed to be larger compared to the permeability assessment of the substratum due to a larger scale. Furthermore the dominating runoff generation processes are more heterogeneously distributed compared to the permeability of the substratum. Both the permeability of the substratum and the dominating runoff generation processes will be derived as percentages of total basin areas in a GIS.

The objective of the study is to compare the information carried by the simplified permeability of the substratum, the dominating runoff generating processes and a combination of both with respect to the winter storm flow coefficient. For this purpose three models that are based on regression analysis will be used. Model results will be assessed with cross-validation, two non-parametric statistical tests and a comparison of their residuals. The use of the permeability of the substratum as a parameter in a regression model may open possibilities for predictions in un-gauged basins concerning their runoff coefficient.

2 Study area

The study area comprises 71 basins located throughout the Grand Duchy of Luxembourg and the Rhineland Palatinate (Germany) and it focuses on the Nahe basin (4011 km², located in the Rhineland Palatinate) and its 16 sub-basins, which are listed in Table 1. The study focuses on the Nahe basin and its sub-basins, which were chosen for the availability of a preliminary GIS-based map that provided dominating runoff generating processes as derived by Steinrücken et al. (2006). All basins have daily discharge measurements for a period of 30 years (1972–2002). Altitudes range from 84 m a.s.l. at the lowest point of the Rhine valley to 817 m a.s.l. on the Hunsrück middle mountain region. The study area has an oceanic temperate climate in the West transforming to a semi-oceanic climate to the East. The temperate humid climate is influenced by the Atlantic Ocean. The macro relief influences rainfall patterns as well. The average annual precipitation ranges from approximately 540 mm/y in the middle part of the study area (Rhine valley) to approximately 1100 mm/y on the higher ridges, with an average annual precipitation of 820 mm/y for the entire study area. The study area is located mainly in the Rhenish Massif and consists largely of schist, siltstone, sandstone and quartzite of Devonian age. The northeastern part is characterized by tectonic dissections of geological strata, hence displaying a heterogeneous geology in comparison to the remainder of the study area (Sauer et al., 2002). The southeastern part of the study area (Pfalz and Rhine valley) consists of an alternation of sandstone, conglomerates and clay of Buntsandstein and of Tertiary sandy, silty deposits and Quaternary Rhine terraces. The overall land use of the study area is 4% urban area, 28% cropland, 22% grassland and 46% forest. However, land use percentages vary between the meso-

Table 1. The percentages of the permeability of the substratum and of the dominant runoff-producing processes of the Nahe basin (outlet at Grolsheim) and its 16 sub-basins.

	Parameters*									
Basin name	L_p [%]	L_{sp} [%]	L_{imp} [%]	D_{SOF1} [%]	D_{SOF2} [%]	D_{SOF3} [%]	D_{SSF1} [%]	D_{SSF2} [%]	D_{SSF3} [%]	D_{DP} [%]
Altenbamberg	68	0	32	9	5	26	0	14	16	31
Boos	48	2	50	8	6	21	5	12	16	31
Enzweiler	32	0	68	7	4	19	4	18	31	19
Eschenau	47	6	47	9	7	21	6	10	14	34
Gensingen	61	0	39	12	3	18	0	6	10	51
Grolsheim	52	2	46	9	5	21	5	11	15	33
Heddesheim	40	0	60	6	5	12	16	7	20	34
Imsweiler	66	0	34	8	6	19	1	10	15	40
Kallenfels	26	0	74	6	5	20	11	10	13	35
Kellenbach	38	0	62	7	5	21	15	6	9	38
Kronweiler	28	0	72	7	3	20	1	15	29	25
Nanzdietschweiler	53	6	41	8	7	18	11	4	10	42
Obermoschel	47	0	53	7	5	34	0	17	13	25
Odenbach	35	0	65	7	5	40	0	19	10	19
Odenbach Glan	52	5	43	9	7	21	3	11	14	35
Steinbach	68	0	32	5	6	15	19	3	9	43
Untersulzbach	83	0	17	9	10	10	1	4	12	55

* L_p: permeable substratum; L_{sp}: semi-permeable substratum; L_{imp}: impermeable substratum; $D_{SOF1,2,3}$: Saturated Overland Flow 1, 2, 3; $D_{SSF1,2,3}$: SubSurface Flow 1, 2, 3; D_{DP}: Deep Percolation

scale basins. For the 71 basins, daily discharge series were available from 1972 until 2002. Rainfall for the same period was obtained from 54 meteorological stations located throughout the study area. In Fig. 1 the permeability of the study area is given. In Fig. 2 the permeability and the dominating runoff generation processes of the Nahe basin and its 16 sub-basins are given. In Table 1 the percentages of these parameters of the Nahe basins are given.

3 Methodology

As dependent variable in the regression models, the winter storm flow coefficient or C-value will be used. The calculation of the C-values can be summarized as follows: firstly, calculate the storm flow of basins by using a base flow separation technique; secondly, build double mass curves of winter storm flow and winter rainfall for each basin and thirdly, calculate the slope in the double mass curve, which denotes the basin specific winter runoff coefficient C. An extensive description of the derivation of the C-value can be found in Pfister et al. (2002), who used the Grand Duchy of Luxembourg as study area. As stated in the introduction, the C-value is supposed to be basin specific, have a strong seasonal variability and should be more or less constant during winter, expressing the saturated state of the basin (Pfister et al., 2002). Since measured discharge during winter is used to calculate the C-values during winter, snowmelt is indirectly taken into account as well in the C-value.

The current study focuses on relationships between C-values and permeability on the one hand and C-values and

dominating runoff generation processes on the other hand. The derivation and comparison of the models can be described as follows:

1. Derivation of three regression models: I, II and III. The models take the C-values of the Nahe basin and its 16 sub-basins as a dependent variable and:

   (a) Model I takes the percentage of the impermeability of the substratum of the Nahe basins as an independent variable. To underpin this relationship, 71 basins located throughout the Grand Duchy of Luxembourg and Rhineland Palatinate, including the Nahe basin and its sub-basins, will be used in a linear regression with their C-values as dependent and permeability of their substrata as independent variables. Since other basin descriptors could also serve as possible regressors (Merz et al. (2006) studied the spatio-temporal variability of event runoff coefficients in Austria and observed a high correlation between values of the runoff coefficient and mean annual precipitation), mean annual precipitation (M_{AP}), mean basin altitude (H_m) and basin area (A_b) are included as well in the regression analysis.

   (b) Model II takes the percentage of one dominating runoff generation process or the percentage of a combination of dominating runoff generation processes of the Nahe basins as an independent variable. A Principal Component Analysis (PCA) will be used to determine the process or combination
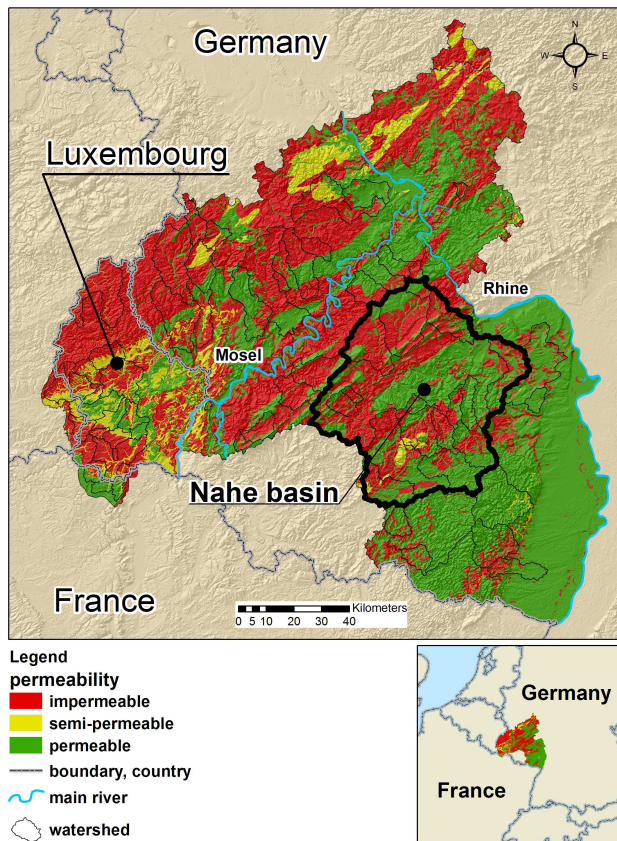
Fig. 1. Permeability map of the Rhineland Palatinate and the Grand Duchy of Luxembourg, based on the GÜK 300.
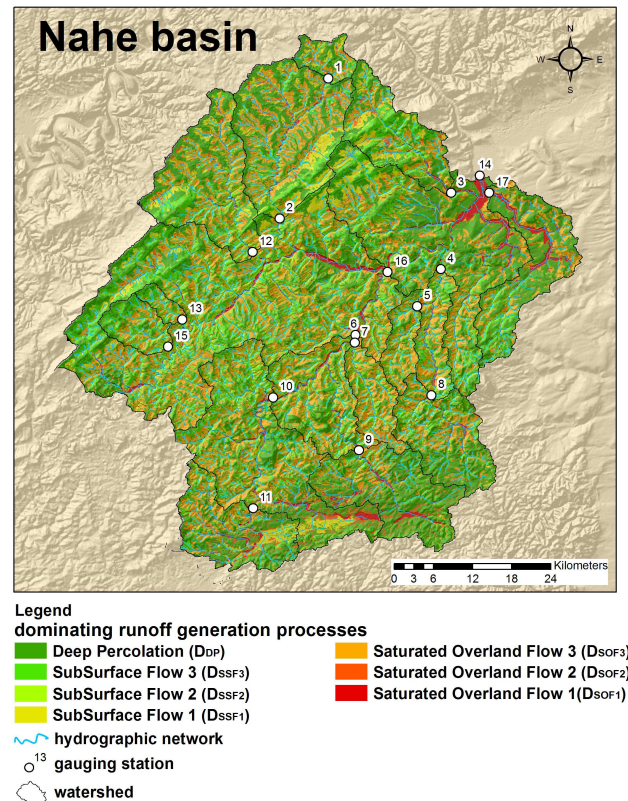

Fig. 2. The preliminary defined dominating runoff production processes of the Nahe basin and its 16 sub-basins (after Steinrücken et al., 2006). For the basin numbers see Table 1.

of processes that bears the highest correlation with the C-value. A PCA is a multivariate technique that produces a set of components (variables) called principal components, which are weighted linear contributions of the original variables (Chatfield and Collins, 1980; James and McCulloch, 1990). In this case the original variables are the C-value and 126 of the possible 127 combinations of the seven dominating runoff generation processes (1 combination combines all seven processes and is therefore redundant). Since the sum of the processes is necessarily 1, only 63 out of the 126 processes are independent. In this study only the positively correlated combinations will be used in the model exercise. However, all 126 combinations have to be used in the PCA. Therefore, it has to be noticed that this renders the negatively correlated combinations as dependent. M_{AP}, H_m and A_b are included as well in the regression analysis.

(c) Model III takes the dominating runoff producing processes and the impermeability of the substratum as independent parameters and is based on a multiple regression. M_{AP}, H_m and A_b are included

in the regression if they appear to be significant in model I and/or model II. In order to obtain the most relevant combination, the five best corresponding combinations of dominating runoff generation processes as derived in step 2, will each be used in a separate multiple regression. The significance of the two parameters, which will be finally used in model III, will be tested with the non-parametric Mann-Whitney U test.

2. Comparing model performances

(a) The performance of the models will be determined with cross-validation, using the RMSE as a comparator value.

(b) The non-parametric Kruskal-Wallis H test (Kruskal and Wallis, 1952) will be used to decide if there is a significant difference between the derived regression models.

(c) The residuals of the models will be compared in order to determine internal mutual differences.

The preliminary models I–III are given in Eqs. (1–3):

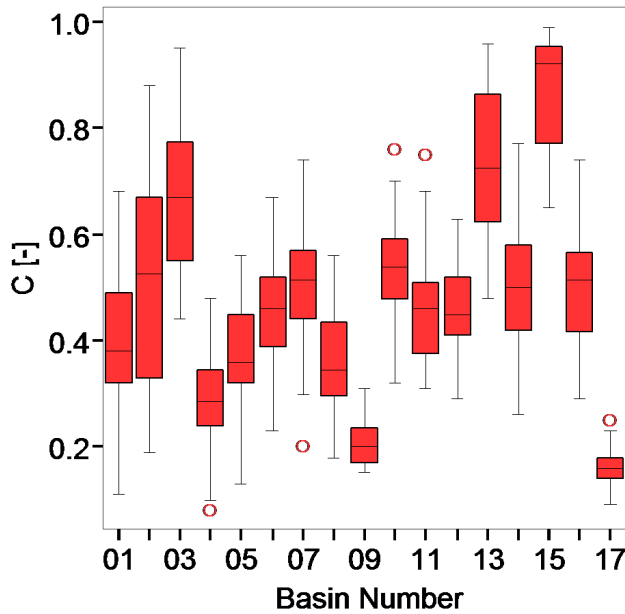$$C_I = a \times L_{imp} + b \times O + r_1 \tag{1}$$

Fig. 3. Box plots of the C-values of the 17 basins (numbers refer to Table 2).

$$C_{II} = c \times D_x + d \times O + r_2 \qquad (2)$$

$$C_{III} = e \times L_{imp} + f \times D_x + g \times O + r_3 \qquad (3)$$

Where:

- C_I, C_{II}, C_{III} are the modeled runoff coefficients of a basin [–]

- a, b, c, d, e, f, g and $r_{1,2,3}$ are constants [–]

- L_{imp} is the percentage of impermeable substratum of a basin [–]

- D_x is the percentage of the dominating runoff generation process or combination of dominating runoff generation processes mostly linked to the C-value of a basin [–]

- O are other basin descriptors (M_{AP}, H_m and A_b) [mm/y, m, km^2]

4 Results and discussion

The calculated C-values of the basins with their standard deviation are listed in Table 2 together with their mean annual precipitation, mean altitude and basin area. In Fig. 3, the box plots of the C-values are given and as can be seen from this figure the annual variability of the C-value was for most of the basins small. The relation between C-values and the percentage of L_{imp} in the 71 basins of the Rhineland Palatinate and the Grand Duchy of Luxembourg showed a good correlation. It could very well be described as linear with an R^2
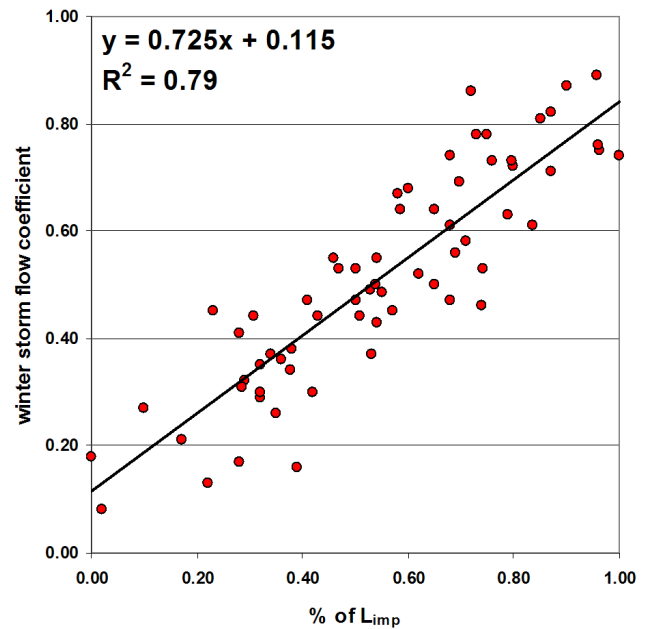

Fig. 4. Correlation between winter storm flow coefficient and percentage of impermeable substratum of Rhineland Palatinate and Luxembourg basins.

of 0.79 (Fig. 4). The residuals did not indicate a bias, thus justifying the relationship. These results were in agreement with the findings of Pfister et al. (2002) for basins located in the Grand Duchy of Luxembourg concerning basin-specific, more or less stable winter storm flow coefficients. Pfister et al. (2002) also found a strong relationship between winter storm flow coefficients and the permeability of the substratum. Apparently, the winter storm flow coefficient appeared to be a good general descriptor of the saturated state of mesoscale basins and well suited to act as a hydrological variable to be used in regionalization procedures.

Table 3 lists the p-values of the multiple regressions with L_{imp} and H_m, L_{imp} and A_b and L_{imp} and M_{AP} as independent variables. Only H_m provided a significant result at the 5% significance level and a non-significant result at the 1% significance level (Table 3) when used as a regressor in a multiple regression with L_{imp}. Although the RMSE was 0.100 when L_{imp} and H_m were used in comparison to a RMSE of 0.114 (see Table 4) when only L_{imp} was used, the RMSEs of the cross-validation were 0.128 when L_{imp} and H_m were used and 0.132 (see also Table 4) when only L_{imp} was used. This latter difference was considered as marginal. The RMSEs of the cross-validation, when A_b and M_{AP} were used in combination with L_{imp} in a multiple regression, turned out to be higher than 0.132. Since, as mentioned before, the study focuses on the interaction between the C-value and L_{imp} on the one hand and the C-value and the dominating runoff processes on the other hand and considering the above-described results it was decided to leave H_m out of the regression.

Table 2. C-values and their standard deviations of the Nahe basin (outlet at Grolsheim) and its 16 sub-basins for a period from 1972 until 2002 and basin area, mean annual precipitation and mean altitude.

Basin name	Basin number [–]	C-value [–]	St dev [–]	Mean annual precipitation [mm/y]	Mean basin altitude [m]	Basin area [km^2]
Altenbamberg	4	0.29	0.09	681	340	318
Boos	16	0.50	0.12	773	471	2833
Enzweiler	13	0.74	0.14	870	493	22.7
Eschenau	10	0.51	0.10	801	397	605
Gensingen	17	0.16	0.06	592	291	197
Grolsheim	14	0.50	0.14	714	449	4011
Heddesheim	3	0.68	0.15	628	395	166
Imsweiler	8	0.37	0.11	676	378	172
Kallenfels	12	0.46	0.11	846	488	253
Kellenbach	2	0.52	0.16	704	439	362
Kronweiler	15	0.86	0.10	790	516	65
Nanzdietschweiler	11	0.47	0.12	976	361	195
Obermoschel	5	0.37	0.11	659	327	62
Odenbach	7	0.5	0.11	841	310	85
Odenbach Glan	6	0.44	0.11	780	380	1069
Steinbach	1	0.3	0.16	685	474	46
Untersulzbach	9	0.21	0.06	736	329	217

Table 3. p-values of the regressors mean basin altitude (H_m), basin area (A_b) and mean annual precipitation (M_AP) when used in a multiple regression with the permeability of the substratum (L_imp) and dominating runoff production processes (Output 1) respectively.

Regressor combinations	p-value	Regressor combinations	p-value
L_imp	0.007	Output 1*	<0.001
H_m	0.053	H_m	0.637
L_imp	<0.001	Output 1	<0.001
A_b	0.407	A_b	0.818
L_imp	0.002	Output 1	<0.001
M_AP	0.352	M_AP	0.226

* Output 1 is: $D_\mathrm{SOF1} + D_\mathrm{SOF2} + D_\mathrm{SSF1} + D_\mathrm{SSF2} + D_\mathrm{SSF3}$

The correlation between C-values and L_imp was less clear for the Nahe basin and its 16 sub-basins than that for the entire study area: an R^2 of only 0.58 was obtained (Fig. 5). The linear regression between the C-values of the Nahe basin and its 16 sub-basins and the percentage of L_imp resulted in model C_I, which is given in Eq. (4). The residuals of model C_I did not indicate a bias.

$$C_\mathrm{I} = 0.865 \times L_\mathrm{imp} + 0.043 \qquad (4)$$

Where: L_imp is the percentage of impermeable substratum of a basin [–].

According to the PCA, the combination of the sum of D_SOF1, D_SOF2, D_SSF1, D_SSF2 and D_SSF3 (Output 1) had the

Fig. 5. Correlation between the percentage of impermeable substratum and the C-value for the Nahe basins and its 16 sub-basins.

strongest correlation with the C-value out of the 63 combinations. Table 3 lists the p-values of the multiple regressions with Output 1 and H_m, Output 1 and A_b and Output 1 and M_AP as independent variables. No significant results at the 5% significance level were obtained. Table 4 lists this combination together with four more combinations that

Table 4. The RMSEs of the performance and cross-validation of the regressions of L_{imp} with H_m, A_b and M_{AP} and Output 1 with H_m, A_b and M_{AP}, the performance and cross-validation of the five combinations of dominating runoff generation processes (drp) that correspond most with the C-value according to the PCA and the performance and cross-validation of the models C_I, C_{II} and C_{III}.

	Best correlated to C-value according to PCA	Model performance	Cross-validation	Cross-validation (drp+L_{imp})
L_{imp} and Output 1 in combination with other basin descriptors		RMSE	RMSE	RMSE
L_{imp} and H_m	–	0.100	0.128	–
L_{imp} and A_b	–	0.111	0.133	–
L_{imp} and M_{AP}	–	0.111	0.139	–
Output 1 and H_m	–	0.086	0.112	–
Output 1 and A_b	–	0.086	0.106	–
Output 1 and M_{AP}	–	0.082	0.163	–
Combination of drp	Output			
$D_{\text{SOF1}} + D_{\text{SOF2}} + D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}$	1	0.086	0.103	0.094
$D_{\text{SOF1}} + D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}$	2	0.088	0.105	0.100
$D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}$	3	0.102	0.101	0.101
$D_{\text{SOF2}} + D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}$	4	0.090	0.103	0.105
$D_{\text{SOF1}} + D_{\text{SSF1}} + D_{\text{SSF2}}$	5	0.141	0.155	0.156
Model				
C_I	–	0.114	0.132	–
C_{II}	–	0.086	0.103	–
C_{III}	–	0.071	0.094	–

corresponded best. As can be observed from Table 4 the sum of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3}, performed best when it was used as a parameter to model the C-value. According to the cross-validation, the combination of the sum of D_{SSF1}, D_{SSF2}, D_{SSF3} (Output 3), was best. The difference between the RMSE of the model performance of Output 1 against Output 3 was markedly larger than the difference between the RMSE of the cross-validation, therefore, in general, Output 1 gave the best performance. When L_{imp} and the separate previously determined five most important dominating hydrological runoff producing processes were used in a multiple regression, Output 1 performed best in the cross-validation (see Table 4). Therefore, Output 1 was chosen as a parameter for the models C_{II} and C_{III}. RMSEs of the cross-validation when H_m, A_b and M_{AP} in combination with Output 1 were used in a multiple regression were higher than the RMSE of the cross-validation when only Output 1 was used (see Table 4). H_m, A_b and M_{AP} were therefore not used in the model.

The correlation between Output 1 and the C-value had an R^2 of 0.76 and was remarkably better than that between L_{imp} and the C-value (Fig. 6). A linear regression between these variables resulted in model C_{II}, which had the C-value as dependent variable and the sum of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3} as an independent variable. The equation of model C_{II} is given in Eq. (5):

$$C_{II} = 2.145 \times [D_{\text{SOF1}} + D_{\text{SOF2}} + D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}] - 0.485 \tag{5}$$

Where: D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3} represent the area percentages of the respective dominating runoff generation processes of a basin [–].

The residuals of model C_{II} did not indicate a bias. Since the C-value can take per definition only values between 0 and 1, extrapolating the winter storm flow coefficient of model C_{II} when the surface area of the dominating runoff-producing processes becomes larger than 60%, becomes problematic (Fig. 6), while this is not the case for model C_I. A multiple regression between the C-value as dependent and L_{imp} and Output 1 as independent variables resulted in model C_{III}. The equation of model C_{III} is given in Eq. (6):

$$C_{III} = 0.408 \times L_{\text{imp}} + 1.559 \\ \times [D_{\text{SOF1}} + D_{\text{SOF2}} + D_{\text{SSF1}} + D_{\text{SSF2}} + D_{\text{SSF3}}] - 0.427 \tag{6}$$

Where:

- D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3} are the area percentages of the respective dominating runoff generation processes of a basin [–]
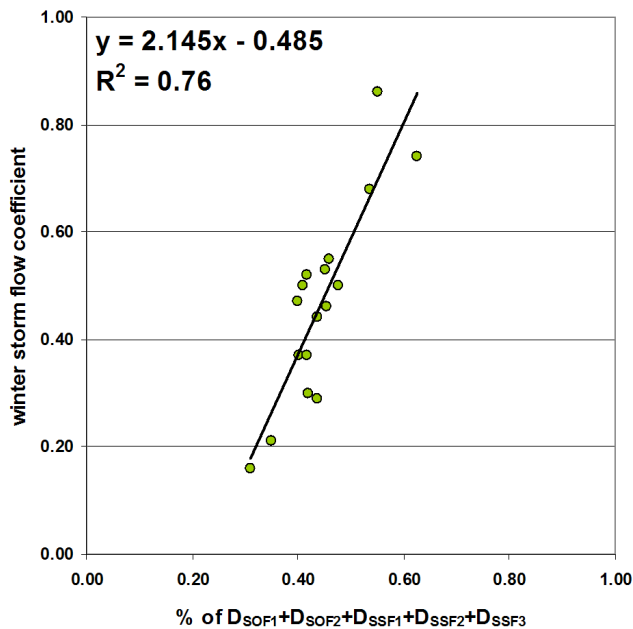
Fig. 6. Correlation between the percentages of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3} and C-value for the Nahe basin and its 16 sub-basins.

Fig. 7. Correlation between the C-values and model C_{III}.

- L_{imp} is the percentage of impermeable substratum of a basin [–]

Figure 7 depicts the correlation between the modeled winter storm flow coefficients by model C_{III} and the calculated winter storm flow coefficients. The residuals of model C_{III} did not indicate a bias.

It should be noticed that four of the five best combinations (listed in Table 4) all carry the sum of the $D_{SSF1,2,3}$ processes in combination with either D_{SOF1} and/or D_{SOF2}. The fifth combination, which is the sum of D_{SOF1}, D_{SSF1} and D_{SSF2} (Output 5 in Table 3), performs considerably less for all performances than the other ones. This indicated that the $D_{SSF1,2,3}$ processes played an important role in determining the C-value during winter. The more or less constant C-value during winter indicated a saturated state of the basins; therefore, presumably a large amount of the soils of the basins should be saturated. The good correlation between the C-values and Output 1, as derived from the PCA, indicated that these processes reflected this saturated state of the basins.

To test the significance of the independent parameters used in model C_{III}, the non-parametric Mann-Whitney U test was applied, which tests the significance of a difference between two samples. The null hypothesis is that the two samples (L_{imp} and sum of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3}) are taken from a common population, so that there should be no consistent difference between the two sets of values. In this case the null hypothesis was rejected with a confidence level of 95%. This means that both parameters (L_{imp} and the sum of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3}) used in
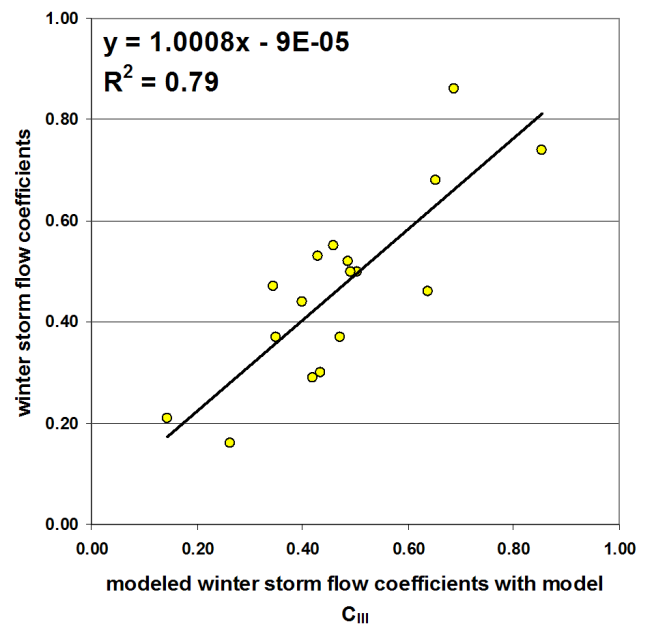
the regression analysis for obtaining model C_{III} came from different populations, hence carrying different information. Model C_{III} did capture the calculation of the C-value best (R^2 of 0.83) in comparison to model C_I (R^2 of 0.58) and C_{II} (R^2 of 0.76) and showed that the combination of L_{imp} and the dominating runoff production processes as derived from the PCA improved the predictability of the C-value.

In order to see if the three models C_I, C_{II} and C_{III} differed substantially from each other (H_0 hypothesis: they are from the same population), the nonparametric Kruskal-Wallis H test (Kruskal and Wallis, 1952) was applied. The test clearly indicated an acceptance of the H_0 hypothesis, with a confidence level of 95%. This means that it can safely be assumed that there is no real difference between the samples (i.e. the modeled values of each model), thus indicating that the three models did possess the same type of information and could be regarded as complementary. In Table 5, the residuals for all three models are given to compare internal mutual differences.

Table 4 lists the RMSEs of the cross-validation. Model C_{III} performed best and model C_I performed worst. The differences between the performances were considerable, especially between model C_I and the models C_{II} and C_{III}. The difference between model C_I and C_{II} could be attributed to a lack of basins to establish a good correlation between the C-value and the permeability of the substratum, since this correlation when the 71 basins were used was rather good. Moreover, the question remains whether with a better assessment of the permeability by using more detailed geological maps a much better performance of model C_I could have been achieved, rendering the use of the dominating runoff

generation processes at this scale as redundant. The gain in model performance for model C_{III} compared to model C_{II} indicated that the parameters of the models C_I and C_{II} did possess some complementary information. It therefore can be argued for that the effort to obtain extra information on the dominating runoff generation processes pays off against a considerable increase in model performance (i.e. performance of model C_I against the performance of model C_{III}).

It turned out that for the models C_I and C_{II} the five worst performing basins were in four cases not the same. This indicated that for these basins the information level between the lithology assessment and the dominating runoff generation processes assessment was apparent. The lithology of the Kallenfels basin, the worst performing basin of model C_I, consists of predominantly claystone and siltstone with inclusion of sandstone (geological formation of the Hunsrückschiefer). In this analysis it was classified as impermeable bedrock (based on the assumption that schist is an impermeable bedrock). Nevertheless, the inclusions of sandstone, which support sandy soils, allow for deep percolation and were not exactly known in this study. Also, a weathered zone or saprolite could have developed at the surface, just beneath the soil, making storage of water possible. The dominating runoff generation processes indicated D_{DP} and D_{SSF3} for large areas of the as L_{imp} classified substratum of the Kallenfels basin. Therefore, this area was probably assessed wrongly in the permeability assessment. The same was the case for the Obermoschel basin. In this basin, various areas consisted of the lower Glan-subgroup, which is an alternation of predominantly grey, partly red clay with inclusions of silt- and sandstone. In the permeability assessment these areas have been classified as L_{imp}, but could well have been assessed as permeable due to their sandstone inclusions. Large areas of D_{DP} and D_{SSF3} indicated by the dominating runoff generation processes again reflected this. A rise of their percentage of permeable area could therefore well be argued for. For the least performing basins of model C_{II}, similar explanations for the assessment of the dominating runoff generation processes as for the assessment of the permeability could be found, which means that the dominating runoff generation processes were assessed as being too large or too small. Since the assessment of the dominating runoff generation processes stops at maximum 2 meters below surface level, deeper lying impermeable substrata are not always taken into account when assessing them. This means that a permeable soil on top of deeper lying impermeable bedrock could still result in subsurface flow as defined by Scherrer and Naef (Scherrer and Naef, 2003).

5 Conclusions

Simple linear models performed rather well to describe runoff-producing processes during winter at the meso-scale. The winter storm flow coefficient could be used as a de-

Table 5. Residuals of the three Nahe models: the five worst performing basins of each model are given in bold.

Basin name	Residuals model C_I [–]	Residuals model C_{II} [–]	Residuals model C_{III} [–]
Altenbamberg	−0.030	**−0.164**	**−0.096**
Boos	0.025	−0.037	−0.020
Enzweiler	0.109	**−0.117**	**−0.086**
Eschenau	0.080	0.049	0.063
Gensingen	**−0.220**	−0.020	−0.055
Grolsheim	0.109	0.052	0.075
Heddesheim	**0.118**	0.017	0.028
Imsweiler	0.033	−0.009	0.030
Kallenfels	**−0.223**	−0.027	**−0.121**
Kellenbach	−0.059	**0.108**	0.042
Kronweiler	**0.194**	**0.165**	**0.136**
Nanzdietschweiler	0.072	0.097	**0.106**
Obermoschel	**−0.131**	−0.038	−0.069
Odenbach	−0.105	0.106	0.023
Odenbach Glan	0.025	−0.010	0.012
Steinbach	−0.020	**-0.116**	−0.058
Untersultzbach	0.020	−0.055	0.022

pendent parameter in a regression analysis. Model performance using a cross-validated RMSE indicated that the simplest model with only one simplified independent parameter (i.e. the permeability of the substratum) performed less well than the model that took into account a more complex independent parameter (i.e. the sum of D_{SOF1}, D_{SOF2}, D_{SSF1}, D_{SSF2} and D_{SSF3}). However, when a model based on multiple regression was used that combined both parameters, the performance of this model was best. The Mann-Whitney U test that was applied to test if the parameters came from different populations and therefore could be used in a multiple regression, gave a negative result.

The non-parametric Kruskal and Wallis H test (Kruskal and Wallis, 1952) that was applied to test if the models agree substantially resulted in an acceptance of the null hypothesis, which proved that the models came form the same population, thus carrying the same type of information. This leads to the conclusion that the permeability of the substratum and the dominating runoff generation processes are complementary bearers of information. Furthermore, comparison of the residuals of the models indicated that badly modeled basins by using the permeability of the substratum as an independent parameter were explained by a lack of information in the permeability of those basins, which could be provided by the preliminary determined dominating runoff generation processes. Badly modeled basins by using a linear combination of these latter processes as an independent parameter were partly explained by a lack of information in the assessment of the processes. This information could be provided by the permeability. As a consequence, the third model that combined both permeability and the dominating runoff generation processes performed better than the other models. To ob-

tain extra information on dominating runoff generation processes paid off against a considerable increase in model performance. However, with a better assessment of the permeability by using more detailed geological maps, a much better performance of the model that used only the permeability of the substratum might have been achieved. Using the permeability as a linear estimator for the C-value in combination with dominating runoff generation processes could determine the winter storm flow coefficient and thereby runoff production areas very well. The impermeability of the substratum may perhaps be used as a parameter for predictions in un-gauged basins. The use of the dominating runoff generation processes though, is due to its modeled nature, still limited to areas with sufficient information on the soils. With the use of GIS-based methods to determine soils (Dobos et al., 2000; Gaddas, 2001) the dominating runoff generation processes could possibly be derived from GIS-based data as well and thereby included for the prediction in un-gauged basins. Testing this assumption with better geological maps and for other regions with a different climate and landscape remains the objective of further study.

References

Blöschl, G.: Scale and scaling in hydrology. Wiener Mitteilungen, Wasser – Abwasser – Gewässer, Wien, Österreich, p. 162, 1996.

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

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, J. Hydrol., 202, 212–230, 1997.

Chatfield, C. and Collins, A. J.: Introduction to multivariate analysis, Chapman & Hall, London, 1980.

Croke, B. and Norton, J.: Regionalisation of rainfall-runoff models, Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs, 3, 1201–1207, 2004.

Dobos, E., Micheli, E., Baumgardner, M. F., Biehl, L., and Helt, T.: Use of combined digital elevation model and satellite radiometric data for regional soil mapping, Geoderma, 97, 367–391, 2000.

Faeh, A. O.: Understanding the processes of discharge formation under extreme precipitation; A study based on the numerical simulation of hillslope experiments, Mitteilung der Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie, ETH Zürich, 150 pp., 1997.

Gaddas, F.: Proposition d'une méthode de cartographie des pédopaysages. Application à la Moyenne Vallée du Rhône, Ph.D. thesis, Institut National Agronomique Paris-Grinon, France, 195 pp., 2001.

James, F. C. and McCulloch, C. E.: Multivariate analysis in ecology and systematics: panacea or Pandora's box?, Annu. Rev. Ecol. Syst., 21, 129–166, 1990.

Kokkonen, T. S., Jakeman, J., Young, P. C., and Koivusalo, H. J.: Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina, Hydrol. Process., 17, 2219–2238, 2003.

Kruskal, W. and Wallis, A.: Use of ranks in one-criterion variance analysis, J. Am. Statist. Assoc., 47, 583–621, 1952.

Mazvimavi, D.: Estimation of Flow Characteristics of Ungauged Catchments: a case study in Zimbabwe, Ph.D. thesis, Wageningen University, The Netherlands, 176 pp., 2003.

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

Merz, R., Blöschl, G., and Parajka, J.: Spatio-temporal variability of event runoff coefficients, J. Hydrol., 331, 591–604, 2006.

Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalization methods for catchment model parameters, Hydrol. Earth Syst. Sci., 10, 353–368, 2006,
http://www.hydrol-earth-syst-sci.net/10/353/2006/.

Pfister, L., Iffly, J.-F., Hoffmann, L., and Humbert, J.: Use of regionalized stormflow coefficients with a view to hydroclimatological hazard mapping, Hydrol. Sci. J., 47, 479–491, 2002.

Post, D. A. and Jakeman, A. J.: Predicting the daily streamflow of ungauged catchments in S.E. Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model, Ecol. Model., 123, 91–104, 1999.

Sauer, D., Scholten, T., Spies, E.-D., and Felix-Henningsen, P.: Pleistocene periglacial slope deposits influenced by geology and relief in the Rhenish Massif, Paper presented at the 17th World Congress of Soil Science, Thailand, 14–21, 2002.

Scherrer, S.: Abflussbildung bei Starkniederschlägen, Identifikation von abflussprozessen mittels künstlicher Niederschläge, VAW – Versuchsanstalt für Wasserbau, Hydrologie und Glaziologie der ETH Zürich, Zrich, 147 pp., 1997.

Scherrer, S. and Naef, F.: A decision scheme to identify dominant flow processes at the plot-scale for the evaluation of contributing areas at the catchments-scale, Hydrol. Process., 17(2), 391–401, 2003.

Schmocker-Fackel, P., Naef, F., and Scherrer, S.: Identifying runoff processes on the plot and catchment scale, Hydrol. Earth Syst. Sci., 11, 891–906, 2007,
http://www.hydrol-earth-syst-sci.net/11/891/2007/.

Steinrücken, U., Behrens, T., and Scholten, T.: Nutzungsbezogene bodenhydrologische Karte: das Einzugsgebiet der Nahe und südlich angrenzende Bereiche (Soilution GbR.), 2006.

Uhlenbrook, S., Roser, S., and Tilch, N.: Hydrological process representation at the meso-scale: the potential of a distributed, conceptual catchment model, J. Hydrol., 291, 278–296, 2004.

Zumstein, J. F., Gille, E., Decloux, J. P., and Paris, P.: Atlas de la lithologie et de la perméabilité du bassin Rhin-Meuse, Agence de l'Eau Rhin-Meuse, Moulin-lès-Metz, France, 1989.