Hydrology and Earth System Sciences

Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium

S. Ly[1], C. Charles[2], and A. Degré[1]

[1]University of Liege, Gembloux Agro-Bio Tech, Unit of Hydrology and Agricultural Hydraulic, Gembloux, Belgium
[2]University of Liege, Gembloux Agro-Bio Tech, Unit of Applied Statistics, Computer Science and Mathematics, Gembloux, Belgium

Abstract. Spatial interpolation of precipitation data is of great importance for hydrological modelling. Geostatistical methods (kriging) are widely applied in spatial interpolation from point measurement to continuous surfaces. The first step in kriging computation is the semi-variogram modelling which usually used only one variogram model for all-moment data. The objective of this paper was to develop different algorithms of spatial interpolation for daily rainfall on $1 \, \text{km}^2$ regular grids in the catchment area and to compare the results of geostatistical and deterministic approaches. This study leaned on 30-yr daily rainfall data of 70 raingages in the hilly landscape of the Ourthe and Ambleve catchments in Belgium ($2908 \, \text{km}^2$). This area lies between 35 and 693 m in elevation and consists of river networks, which are tributaries of the Meuse River. For geostatistical algorithms, seven semi-variogram models (logarithmic, power, exponential, Gaussian, rational quadratic, spherical and penta-spherical) were fitted to daily sample semi-variogram on a daily basis. These seven variogram models were also adopted to avoid negative interpolated rainfall. The elevation, extracted from a digital elevation model, was incorporated into multivariate geostatistics. Seven validation raingages and cross validation were used to compare the interpolation performance of these algorithms applied to different densities of raingages. We found that between the seven variogram models used, the Gaussian model was the most frequently best fit. Using seven variogram models can avoid negative daily rainfall in ordinary kriging. The negative estimates of kriging were observed for convective more than stratiform rain. The performance of the different methods varied slightly according to the density of raingages, partic-

ularly between 8 and 70 raingages but it was much different for interpolation using 4 raingages. Spatial interpolation with the geostatistical and Inverse Distance Weighting (IDW) algorithms outperformed considerably the interpolation with the Thiessen polygon, commonly used in various hydrological models. Integrating elevation into Kriging with an External Drift (KED) and Ordinary Cokriging (OCK) did not improve the interpolation accuracy for daily rainfall. Ordinary Kriging (ORK) and IDW were considered to be the best methods, as they provided smallest RMSE value for nearly all cases. Care should be taken in applying UNK and KED when interpolating daily rainfall with very few neighbourhood sample points. These recommendations complement the results reported in the literature. ORK, UNK and KED using only spherical model offered a slightly better result whereas OCK using seven variogram models achieved better result.

1 Introduction

Basin management, including hydrological and water quality applications, requires data on the very important precipitation parameter. These data are often collected using raingages, and hence they are point data. However, the use of a single raingage as rainfall input carries great uncertainties regarding runoff estimation (Faurès et al., 1995 and Chaubey et al., 1999). This presents a great problem for the prediction of discharge, groundwater level and soil moisture, especially if the raingage is located outside the catchment (Schuurmans and Bierkens, 2007). As a result, some applications such as rainfall erosivity mapping (Aronica and Ferro, 1997; Goovaerts, 1999; Hoyos et al., 2005; Nyssen et al., 2005; Angulo-Martínez and Beguería, 2009; Angulo-Martínez et

Correspondence to: S. Ly
(sarann.ly@doct.ulg.ac.be)

al., 2009) and hydrological modelling (Syed et al., 2003; Kobold and Sušelj, 2005; Gabellani et al., 2007; Cole and Moore, 2008; Collischonn, et al., 2008; Ruelland et al., 2008; Moulin et al., 2009) require rainfall data that are spatially continuous. The quality of such result is thus determined by the quality of the continuous spatial rainfall (Singh, 1997; Andréassian et al., 2001; Kobold and Sušelj, 2005; Leander et al., 2008; Moulin et al., 2009).

The generation of continuous surfaces starting from irregularly distributed data is a task for many disciplines. It can be performed by a variety of methods but the difficulty lies in choosing the one that best reproduces the actual surface (Caruso and Quarta, 1998). Regarding the rainfall, indirect estimates of continuous surface based on the measurement of related ancillary variables have been provided since the late 1960s by ground-based meteorological RADARs and by remote sensing devices carried on satellite platforms. The significance and reliability of such indirect methods for hydrological purposes have yet to be determined. The methods must be calibrated and validated using historical data (Lanza et al., 2001). However for direct ground-based measurement, spatial interpolation techniques can be used and broadly classified into two main groups: *deterministic* and *geostatistical*. The most frequently used deterministic methods in spatial interpolation are the Thiessen polygon and Inverse Distance Weighting (IDW). The geostatistical method constitutes a discipline involving mathematics and earth sciences. A South African mine engineer, Krige, is a precursor of geostatistics. Nevertheless, the term "kriging" and the formalism of this method are coined by Matheron (1971).

Regarding the spatial interpolation of rainfall, Dirks et al. (1998) compared Inverse Distance Weighting, the Thiessen polygon and kriging in interpolating rainfall data from a network of thirteen raingages on Norfolk Island. They recommended the use of IDW for interpolations for spatially dense networks. Nalder and Wein (1998) used cross validation to evaluate four forms of kriging and three simple alternatives for spatial interpolation of climatic data. They found that IDW had a smaller error of estimates than Ordinary Kriging (ORK) and Universal Kriging (UNK) in interpolating monthly precipitation in the Canadian boreal forest. Buytaert et al. (2006) studied the variability of spatial and temporal rainfall in the south Ecuadorian Andes using the Thiessen polygon and kriging. Their study suggested that spatial interpolation with kriging gives better a result than Thiessen polygon, and the accuracy of both methods improves when external trends are incorporated. Basistha et al. (2008) analysed the spatial distribution of rainfall in the Indian Himalayas using both deterministic and geostatistical methods. They reported that UNK was the most suitable method, followed by ORK and IDW.

In geostatistical methods, there are several possibilities to incorporate secondary data to improve primary data. Since modern equipments such radar, microwave links, satellites, etc., are available for some countries, the products from those equipments are usually used to improve rainfall interpolation. Berne et al. (2004) analysed the temporal and spatial structure of rainfall using high resolution raingage and radar measurement. Schuurmans et al. (2007) used range-corrected daily radar composites to integrate into two methods of geostatistics, and it proved to be more accurate than a method that uses rainfall alone. Velasco-Forero et al. (2009) and Schiemann et al. (2011) also integrated radar data into two geostatistical methods but they focused on using an automatic non-parametric variogram and applied to hourly rainfall. Nevertheless, in the area where the modern equipments are not available, elevation, especially extracted from a digital elevation model (DEM), are a cheaper and more widely available data which can be used to incorporate into multivariate geostatistics of rainfall. Goovaerts (2000) and Lloyd (2005) used elevation as secondary data to incorporate into multivariate geostatistics for monthly and annual rainfall and compared these results with those of deterministic methods. Goovaerts (2000) found that inverse square distance and Thiessen polygon give larger prediction errors than the three multivariate geostatistical algorithms. Lloyd (2005) concluded that Kriging with External Drift (KED) provides the most accurate estimates of precipitation for all months from March to December whereas for January and February ORK provided the most accurate estimates, involving with monthly precipitation mapping in Great Britain from sparse data. Verworn and Haberlandt (2011) recommended the use of additional information from topography for KED. They applied to hourly rainfall with a high spatial resolution. However, the studies of multivariate geostatistics using elevation for daily rainfall are rather limited.

Most of applications considered only monthly or annual time steps for spatial interpolation of precipitation (Hevesi et al., 1992; Goovaerts, 2000; Boer et al., 2001; Todini, 2001; Marquínez et al., 2003; Vicente-Serrano et al., 2003; Lloyd, 2005) while some others used hourly time steps for large-scale extreme rainfall events (Haberlandt, 2007). Most recently, hourly precipitation was spatially interpolated with multivariate geostatistical method in northern part of Germany (Verworn and Haberlandt, 2011). In addition, hydrological applications for urban catchments in Mediterranean regions require a temporal resolution of about 5 min for an area of 1000 hectares and 3 min for an area of 100 hectares (Berne et al., 2004). Little experience exists on the use of geostatistics for daily rainfall at catchment scale (notable exceptions include Beek, 1992, who used only four days, Schuurmans et al., 2007, who restricted to radar data and some events and Carrera-Hemández and Gaskin, 2007, who limited to only two-month periods). The reliability of predictions may change if a different time step is chosen. Daily rainfall is a major meteorological input to water resources and agricultural modelling systems since rainfall data are mostly available in daily time step in national or regional measurement. The question is whether the best method applied to monthly or annual rainfall is applicable to daily

rainfall since precipitation pattern differences between daily and monthly timescales (Johnson and Hanson, 1995). Moreover, the analysis of the spatial distribution of daily rainfall is difficult mainly because of intermittence and large variability (Carrera-Hemández and Gaskin, 2007). Current practice, attention will be thus focused on the applications to the long series of daily rainfall at the catchment scale.

Very few studies, if at all, attempt to analyse the effect of raingage density used for interpolation. Borga and Vizzaccaro (1996) compared kriging and multiquadratic surface fitting to interpolate hourly rainfall and found that kriging performs better at lower raingage density while at higher raingage density the accuracy of both interpolators is similar. Campling et al. (2001) used ordinary kriging with unscaled variogram model to optimize the number and location of raingages. Spadavecchia and Williams (2009) varied the number of interpolation data used in both space and time and assessed its impact on interpolation skill of spatio-temporal geostatistical methods. The applications to different densities of raingages were restricted to only one or two methods. This might be because it was very cumbersome in terms of computation time. However, the comparison between more techniques as regards to raingage density may provide some insights in terms of particular strengths, weaknesses and applicability of the methods since the computation facilities are well developed and available at the present time.

The objective of this study is to obtain the best interpolation method to produce daily rainfall data for hydrological models. We develop algorithms for spatial interpolation of daily rainfall data on $1\,\mathrm{km}^2$ regular grids at the catchment scale, taking into consideration the automatic choice of a theoretical variogram model, that is based on daily data, among the sevens variogram models for geostatistical methods. The algorithms are applied to different cases degenerating into different numbers of raingages inside and surrounding the catchments. Previous studies usually applied the same theoretical variogram model for all time steps. Most of them chose spherical model when using geostatistical methods to interpolate rainfall. Recently, van de Beek et al. (2011) used the spherical model to analyse the seasonal variogram parameters of daily rainfall in the Netherlands. Verworn and Haberlandt (2011) also inferred spherical model for spatial interpolation of hourly rainfall in northern part of Germany. However, kriging can lead to negative estimates (Deutsch, 1996). Negative weights in ordinary kriging occur when data close to the location being estimated display outlying data. Depending on the variogram model and the data values, the negative weights can be significant. Also negative weights when applied to high data values may lead to negative and nonphysical estimates. Generally, there are two ways to avoid negative value: a posteriori correction by Deutsch (1996) or replace all negative interpolated values with a zero value. Both are pragmatic solution but not ideal. For some cases, the negative values are significant and it is not really fair to replace them by zero. The tackle to the solution of the neg-

ative weights are very limited in the application to rainfall (exception for Haberlandt, 2007, who set negative rainfall estimates to zero). Therefore, we try to change the variogram model from one another to avoid negative rainfall.

Based on the literature reviewed above, the paper has several innovative aspects. Very few studies focused on using elevation to integrate into multivariate geostatistics for spatial interpolation of daily rainfall. Daily rainfall has a particular stochastic nature which differs from monthly rainfall (Johnson and Hanson, 1995). Most recent studies used radar rainfall in daily time step. Therefore, it is appealing to explore whether incorporation of elevation as auxiliary variable improves interpolation result since rainfall data are mostly available in a daily time step. Analyses of the effect of raingage density on interpolation methods have been relatively little studied. Some studies focused mainly on the effect within only one method (ORK, UNK or KED) or two methods (multiquadratic surface fitting and kriging). However, our study focuses on the comparison between six different techniques. One specific contribution of the paper illustrates how ordinary kriging can get rid of negative result using seven variogram models. Such analyses in this study can be valuable for engineers, hydrologists or decision makers working with sparse raingage data.

2 Materials and methods

In this paper, geostatistical algorithms (ordinary kriging, universal kriging, kriging with an external drift and ordinary cokriging), deterministic algorithms (Thiessen polygon and inverse distance weighting) were developed using Fortran 90 to produce the daily rainfall of each grid from 1976 to 2005. The performance of these methods was then evaluated. For some days when zero rainfall occurred at all raingages, we supposed that there was no rainfall in the whole catchments, and thus no interpolation was made for these days.

2.1 Interpolation procedures

The interpolation methods used in this paper will be briefly introduced. A detailed presentation of geostatistical theories can be found in Cressie (1991), Goovaerts (1997), Chilès and Delfiner (1999), and Webster and Oliver (2007).

Spatial interpolation is generally carried out by estimating a regionalized value at un-sampled points from a weight of observed regionalized values. In this study, un-sampled points refer to the centres of 1-km^2-regular grids in the catchment area. The general formula for spatial interpolation is as follows:

$$Zg = \sum_{i=1}^{ns} \lambda_i Zs_i \tag{1}$$

where Zg is the interpolated value at point g, Zs_i is the observed value at point i, ns is the total number of observed

points (raingages) and $\lambda = \lambda_i$ is the weight contributing to the interpolation.

The problem lies in calculating the weights λ which will be used in the interpolation. The different methods for computing the weights will be presented in the following sections.

2.1.1 Deterministic methods

Thiessen polygon (THI)

For the Thiessen polygon, the catchment area is divided into polygons so that each polygon contains a single point of sampling (Chow, 1964). Each interpolated point (centre of each grid) takes the value of the closest sampled point. The advantage of this method is its simplicity. However, the disadvantages of this method are obvious – the estimation is based on only one measurement and the information on neighbouring points is ignored. Moreover, there are sudden jumps of discontinuity in the passage from one polygon to another.

Inverse Distance Weighting (IDW)

Inverse Distance Weighting (IDW) estimates values at unsampled points by the weighted average of observed data at surrounding points. So, this can be defined as a distance reverse function of each point from neighbouring points (Teegavarupu and Chandramouli, 2005). That means by using a linear combination of values at a known sampled point, values at un-sampled points can be calculated. IDW relies on the theory that the unknown value of a point is more influenced by closer points than by points further away. The weight can be computed by:

$$\lambda_i = \frac{\frac{1}{|D_i|^d}}{\sum\limits_{i=1}^{ns} \frac{1}{|D_i|^d}}, \quad d > 0 \tag{2}$$

where D_i is the distance between sampled and un-sampled points. The d parameter is specified as a geometric form for the weight while other specifications are possible. This specification implies that if the power d is larger than 1, the so-called distance-decay effect will be more than proportional to an increase in distance, and vice versa. Thus, small power d tends to give estimated values as averages of Zs_i in the neighbourhood, while large power d tends to give larger weights to the nearest points and increasingly down-weights points further away (Lu and Wong, 2008). Using a power value of 2 for daily and monthly time steps, 3 for hourly and 1 for yearly would appear to minimize the interpolation errors (Dirks et al., 1998). Furthermore, this power d is usually set to 2, following Goovaert (2000) and Lloyd (2005) and hence inverse square distances are used in the estimation. Consequently, a power value of 2 was adopted for IDW in this study.

2.1.2 Geostatistical methods

Geostatistical methods use the semi-variograms as a core tool to characterize the spatial dependence in the property of interest. Figure 1 shows a simplified flowchart of kriging computations procedures carried out in this study.

Variogram modelling

First of all, the experimental semi-variogram was calculated as a half the squares difference between paired values to distance by which they were separated:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (Zs_i - Z(s_i + h))^2 \tag{3}$$

where $N(h)$ is the number of pairs of data locations at distance h apart. The hypotheses of spatial variability were here homogeneity and an isotropic spatial pattern due to the lack of number of point data, and hence identical variability in all directions.

In practice, the average squared distance was obtained for all pairs separated by a range of distances and these average squares differences were plotted against the average separation distance. In this study, the bin size depends on the number of gauges used. A theoretical model might then be fitted to the experimental semi-variogram and the coefficient of this model could be used for kriging. Most previous studies have used only one theoretical model for each time step, and these were mostly in monthly or yearly steps (Hevesi et al., 1992; Goovaerts, 2000; Boer et al., 2001; Todini, 2001; Marquínez et al., 2003; Lloyd, 2005). However, this paper focuses on daily data over 30 yr. On a daily basis, rainfall has different spatial variability. In this study, we dealt with the fitting of the semi-variogram for every day of our 30-yr period. In order to do this, we used seven existing theoretical models, as presented below:

– De Wijs (logarithmic) model:

$$\gamma(h; \theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1 \ln\|h\|, & h \neq 0, \end{cases} \tag{4}$$

for $\theta_0 \geq 0$ and $\theta_1 \geq 0$.

This model is often used for describing variables regularized by a sampling support, which implies that the infinite variance problem disappears (Chilès and Delfiner, 1999; Diggle and Ribeiro Jr., 2007).

– Power model:

$$\gamma(h; \theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1 \|h\|^{\theta_2}, & h \neq 0, \end{cases} \tag{5}$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $0 \leq \theta_2 < 2$.

The parameter θ_2 is restricted between 0 and 2 in order that the model is a conditional positive definite function and then a permissible variogram model (Pardo-Igúzquiza, 1998).
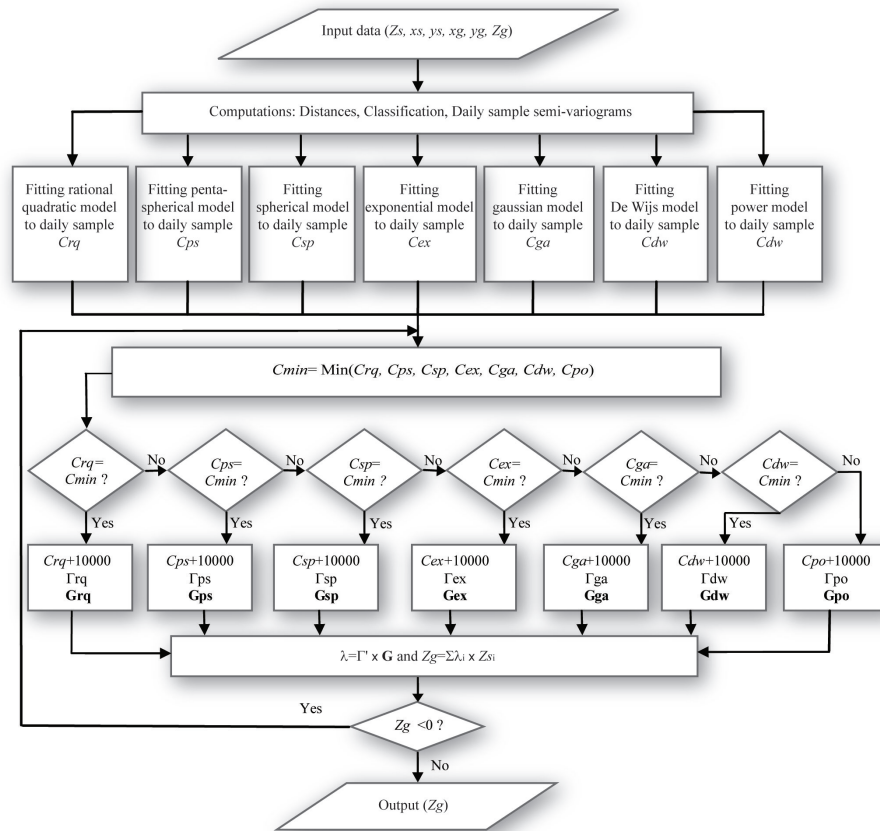
Fig. 1. Flowchart showing the simplified procedure of kriging computation. Cxx refers to weighted least squares criterion of each model fitting (see Eq. 11), Cxx +10 000 refers added value to the criterion to avoid the variogram model from being used again when negative rainfall results occur.

– Exponential model:

$$\gamma(h;\theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1[1 - \exp(-3\|h\|/\theta_2)], & h \neq 0, \end{cases} \quad (6)$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $\theta_2 \geq 0$.

– Gaussian model:

$$\gamma(h;\theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1\left\{1 - \exp\left[-3(\|h\|/\theta_2)^2\right]\right\}, & h \neq 0, \end{cases} \quad (7)$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $\theta_2 \geq 0$.

– Rational quadratic model:

$$\gamma(h;\theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1 \dfrac{19\left(\|h\|/\theta_2\right)^2}{1+19\left(\|h\|/\theta_2\right)^2}, & h \neq 0, \end{cases} \quad (8)$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $\theta_2 \geq 0$.

– Spherical model:

$$\gamma(h;\theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1\left(\dfrac{3\|h\|}{2\theta_2} - \dfrac{1}{2}\left(\dfrac{\|h\|}{\theta_2}\right)^3\right), & 0 < \|h\| \leq \theta_2, \\ \theta_0 + \theta_1, & h > \theta_2, \end{cases} \quad (9)$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $\theta_2 \geq 0$.

– Penta-spherical model

$$\gamma(h;\theta) = \begin{cases} 0, & h = 0, \\ \theta_0 + \theta_1\left(\dfrac{15\|h\|}{8\theta_2} - \dfrac{5}{4}\left(\dfrac{\|h\|}{\theta_2}\right)^3 + \dfrac{3}{8}\left(\dfrac{\|h\|}{\theta_2}\right)^5\right), & 0 < \|h\| \leq \theta_2, \\ \theta_0 + \theta_1, & h > \theta_2, \end{cases} \quad (10)$$

for $\theta_0 \geq 0$, $\theta_1 \geq 0$ and $\theta_2 \geq 0$.

Each of these models was combined with a nugget effect. The most common methods of fitting semi-variogram models to experimental semi-variograms are performed using manual fitting procedures (Nalder and Wein, 1998 and Harberlandt, 2007). However, this is not an appropriate approach because it depends on the expertise and the information in the field. Moreover, this procedure was not feasible for daily data of 30 yr; hence instead an automatic procedure was necessary. Cressie (1985) proposed weighted least squares, used
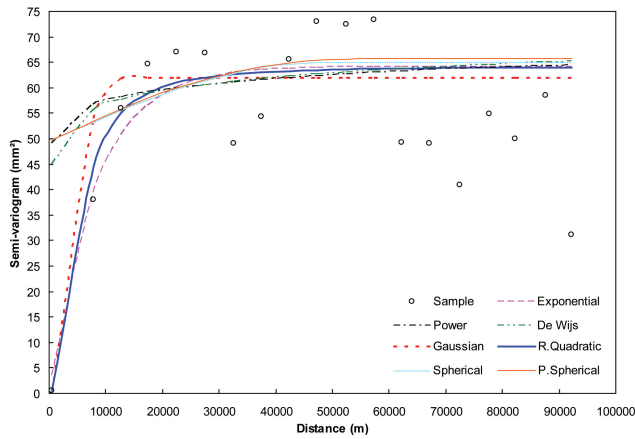
Fig. 2. Sample semi-variogram of daily rainfall (7 August 1991) with seven models fitted: the Rational Quadratic Model is best fitted.

in this study, as a reasonable compromise between the efficiency of generalized least squares and the simplicity of ordinary least squares for fitting a semi-variogram model to an experimental semi-variogram. He introduced an approximation to the weighted least squares criterion, which reduced the estimation problem to iteratively reweighted least squares:

$$C(h;\theta) = \sum_{i=1}^{K} |N(h_i)| \left(\frac{\hat{\gamma}(h_i)}{\gamma(h_i;\theta)} - 1 \right)^2 \qquad (11)$$

where h_1, h_2, ..., h_k are equally spaced lags at which the semi-variogram is estimated. In this study, the coefficients of a model for kriging were chosen when the coefficient C was least among the iteration processes. Moreover, a model was chosen for each day by considering the model which provided the smallest C among the seven models. Figure 2 shows an example of fitting seven variogram models to the sample semi-variogram of rainfall on 7 August 1991. The rational quadratic model was chosen for this day. For this study, the kriging methods were computed separately. If a variogram model yielded negative rainfall somewhere, the variogram model changed only for that kriging method.

The coefficients of the chosen model were then used to determine the weight through equation systems of different types of kriging: Ordinary Kriging (ORK), Universal Kriging (UNK), and Kriging with an External Drift (KED).

Ordinary Kriging (ORK)

The weights are obtained such that the estimation is unbiased and the variance is minimized. The ORK system of (ns+1)

equations, is as follow:

$$\begin{cases} \sum_{i=1}^{ns} \lambda_i \gamma_{ij} - \mu = \gamma_{i0} & \text{for } j=1,\ldots,ns \\ \sum_{i=1}^{ns} \lambda_i = 1 \end{cases} \qquad (12)$$

Where γ_{ij} represents the semi-variances of Zs between locations i and j, and μ is the Lagrange parameter.

This system can be shown in matrix form to facilitate the resolution:

$$\mathbf{\Gamma} \times \boldsymbol{\lambda} = \boldsymbol{G} \implies \boldsymbol{\lambda} = \mathbf{\Gamma}^{-1} \times \boldsymbol{G} \qquad (13)$$

The weights λ_i, obtained through this system are inserted into Eq. (1) to make the prediction. The unbiased estimate is assured by the constraint of the sum of the weight to 1, which requires the definition of the Lagrange parameter.

Universal Kriging (UNK)

This assumes that spatial variation in estimated values has a structural component in addition to the spatial correlation between known points (Basista et al., 2008). Typically, UNK incorporates a trend surface equation in the kriging process. This can be either a first order polynomial or it can be a quadratic surface defined by a second order polynomial. The prediction is computed when the weights are such that the prediction is unbiased and the variance is minimized. The same process as in ORK is followed. The system of (ns+L+1) can be written as:

$$\begin{cases} \sum_{i=1}^{ns} \lambda_i \gamma_{ij} + \sum_{l=1}^{L} \mu_l f_i^l = \gamma_{i0} & \text{for } i=1,\ldots,ns \\ \sum_{i=1}^{ns} \lambda_i f_i^l = f_0^l & \text{for } l=0,\ldots,L \end{cases} \qquad (14)$$

where γ_{ij} represents the semi-variances of Zs between locations i and j, and μ_l are the Lagrange parameters and f is the mean which is a function of spatial coordinates. This study dealt with linear trend, hence $L=2$, $f_i^0 = 1$, $f_i^1 = x_i$, $f_i^2 = y_i$ (x and y are abscise and ordinate of the points). When $L=0$, $\sum_{i=1}^{ns} \lambda_i = 1$ which is the constraint of un-bias. The system can be also written in matrix form (Eq. 13) and the weights λ_i can be computed to make the prediction.

Kriging with an External Drift (KED)

We hypothesized that the variable of interest presents a structure of ensemble modelled by a secondary variable. The spatial behaviour of the secondary variable is similar to an indicator of general trend, the so-called external drift, representative of a representation of predictions, regarding the considered geographical domain. This study dealt with the
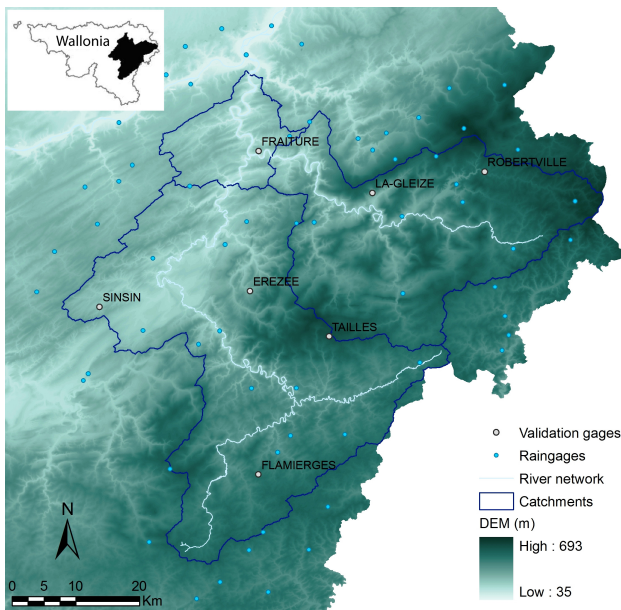
Fig. 3. Location of raingages and Digital Elevation Model of the Ourthe and Ambleve catchments, Walloon, Belgium.

trend model of the combined function of spatial coordinates and elevation as the secondary variable simultaneously.

All variogram models may still be able to produce negative results with UNK and KED at some days. We modified the Kriging system to disallow negative weights, based on the a posteriori correction described in Deutsch (1996).

Ordinary Cokriging (OCK)

The last of the geostatistical methods dealt with in this study is Ordinary Cokriging (OCK), which is another approach to incorporating secondary information in order to improve the spatial interpolation. Goovaerts (2000) stated that using multiple secondary variables can lead to unstable cokriging systems. Thus, only elevation (Ys) was considered in this paper. The cokriging estimate is:

$$Zg = \sum_{i=1}^{ns} \lambda_i Z s_i + \sum_{i=1}^{ns} \alpha_i Y s_i \tag{15}$$

where all secondary variables are obtained at the same points of variable of interest. As with ORK, the objective is to minimize the variance under the constraint of un-bias, which gives a very complex system of ($ns+ns+2$) equations:

$$\begin{cases} \sum_{i=1}^{ns} \lambda_i \gamma(Z_i, Z_j) + \sum_{i=1}^{ns} \alpha_i \gamma(Z_i, Y_j) + \mu_Z = \gamma(Z_0, Z_i) & \text{for } j=1,\ldots,ns \\ \sum_{i=1}^{ns} \lambda_i \gamma(Y_i, Z_j) + \sum_{i=1}^{ns} \alpha_i \gamma(Y_i, Y_j) + \mu_Y = \gamma(Z_0, Y_i) & \text{for } j=1,\ldots,ns \\ \sum_{i=1}^{ns} \lambda_i = 1 \\ \sum_{i=1}^{ns} \alpha_i = 0 \end{cases} \tag{16}$$

The system can also be written in the matrix form. There are two Lagrange parameters to take into account for the constraints on the weight of primary and secondary data. The input information ($\gamma(Z_i, Z_j)$, $\gamma(Y_i, Y_j)$ and $\gamma(Z_i, Y_j)$) represents the values of direct semi-variograms of primary and secondary variables and cross semi-variograms of primary and secondary variables respectively for spaced distances. The experimental cross-semi-variograms were computed as:

$$\hat{\gamma}_{ZY}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Zs_i - Z(s_i+h)][Ys_i - Y(s_i+h)] \tag{17}$$

Modelling the co-regionalization between two variables Z and Y involves choosing and fitting theoretical models to the two direct semi-variograms $\gamma(Z_i, Z_j)$ and $\gamma(Y_i, Y_j)$ plus the cross semi-variogram $\gamma(Z_i, Y_j)$. The difficulty lies in the fact that the three models can not be built independently from one another. The easiest approach consists of modelling the three semi-variograms as linear combinations of the same set of basic semi-variogram models (Goovaerts, 1998).

The coefficients of the fitted models are used to determine the weight through the equation systems of Ordinary Cokriging (Eq. 16).

2.2 Study area, data and case study

In this study, we used daily rainfall data of 30 yr (1976–2005) from 70 raingages within and surrounding the hilly landscape of the Ourthe and Ambleve catchments ($2908 \, \text{km}^2$). These catchments were divided into regular grids of $1 \, \text{km}^2$. The catchment area lies between 35 and 693 m in elevation, and is located in the Ardennes hill range in the south-eastern part of Belgium, called the Walloon region (Fig. 3). The Ourthe River is an important tributary of the Meuse River. Since the higher Condroz region acts as a natural boundary, the Ourthe flows in a northerly direction. Several smaller tributaries, such as the Vesdre and the Ambleve, join the Ourthe River along its way towards Liege, where it eventually joins the river Meuse.

The precipitation data were provided by the Royal Meteorological Institute of Belgium. The elevation data used for this study are extracted from the Digital Elevation Model (DEM) provided by the ERRUISSOL project (Demarcin et al., 2009).

In this study, the interpolation procedures were applied to different case studies. Firstly, we used 70 raingages. Then we

degenerated into 60, 50, 40, 30, 20, 8 and 4 raingages. For variogram fitting, the bin size is 5 km for 70 and 60 raingages used, 10 km for 50, 40 and 30 raingages used, 15 km for 20 and 8 raingages used and 33 km for 4 raingages used.

2.3 Evaluation criteria for interpolators

The evaluation of such a comparison of different interpolators was usually made by cross validation which involves temporarily discarding data from the sample data set; the value at the same location is then estimated using the remaining samples (Isaaks and Srivastava, 1989). Most of authors cited in this paper used a cross validation technique with monthly or annual time steps. The sample size from the cross validation is the number of sample data (number of existing raingages). Nevertheless, it would be time consuming to use cross validation for the daily time steps of 30-yr precipitation. Therefore, two validation approaches were performed.

Firstly, seven raingages in the study area were randomly selected to be used for validation, in view of the fact that the existing observed daily rainfall series of these seven raingages provided a large enough sample size. These seven raingages are FLAMIERGES (elevation 496.88), FRAITURE (elevation 235.54), LA GLEIZE (elevation 333.95), TAILLES (elevation 608.67), ROBERTVILLE (elevation 514.82), EREZEE (elevation 320.87) and SINSIN (elevation 236.1). Moreover, they are spread over the catchment area and also cover the whole elevation's range of the catchments (Fig. 3). These raingages were not included when we used 60, 50, 40, 30, 20, 8 and 4 raingages for interpolation. When using 70 raingages, one of seven raingages was temporarily removed from the 70 sample data set for each computation; the value at the same location was then estimated using the 69 remaining samples. The interpolated rainfalls were then compared to observed time series of daily rainfall at these seven raingages.

Secondly, the cross validations are performed for two periods of very distinct rainfall patterns selected from 30-yr dataset. The first period is in winter (15 to 30 December 1993) where the rainfall is very intense and mostly stratiform, and the second period is in summer (3 to 18 July 1994) where the rainfall presents mostly convective type-variable amount in the catchment area.

In order to determine whether there were large differences in the results when using only one variogram model in the interpolation techniques, we modified our geostatistical algorithms using only one spherical model and simply putting negative rainfall to zero. We chose this model, mainly because it is one of the models most commonly used for interpolating rainfall found in the literature mentioned in this paper. These approaches were also applied to the different densities of raingages.

Root Mean Square Error (RMSE) has been used as a criteria of comparison in many studies related to spatial interpolation of rainfall such as high-resolution studies of rainfall

on Norfolk Island (Dirks et al., 1998), assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain (Lloyd, 2005), comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale in Mountain Appennies chain (Attore et al., 2007), and spatial distribution of rainfall in the Indian Himalayas (Basistha et al., 2008).

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Zs_i^* - Zs_i\right)^2} \qquad (17)$$

Zs^*:   observed value at the raingage;
Zs: interpolated value at the raingage;
n: sample size (total days of data series or total number of available raingages).

The most accurate algorithm has an RMSE value closest to zero. Although all geostatistical methods provide an estimate of the error variance, but this value has not been retained as a performance criterion because it is not adequate to delimit the reliability of kriging estimate (Goovaert, 2000).

3 Results

3.1 Variogram models of daily rainfall

For each day, we generated seven variogram models for all different number of raingages used. Semi-variance increased according to the separation distance, explaining that two rainfall data close to each other were more similar, and hence their squared difference was less significant, than those that are farther apart. For a day of winter period (20 December 1993), the variograms corresponding to the 70 and 30 raingages used, exhibited nearly the same structure (Fig. 4). The spherical models were the best fits. They provided the same range of about 94 km, the nugget effect of 0.289 and 0.23 mm^2, and the sill of about 118.87 and 125.33 mm^2 for the 70 and 30 respective raingages used. For sparse density of raingages, the Gaussian models were the best fits, providing the range of about 66 km and 80 km, the nugget effect of 0.24 and 8.27 mm^2, and the sill of 135 and 141 mm^2 for 8 and 4 respective raingages used.

From 30 yr of daily rainfall, the variogram models changed significantly from day to day. Among the seven models used, the Gaussian model was the most frequently best fitted to the daily sample semi-variogram (Fig. 5a). For best fits, the frequency of the rational quadratic model is very much lower than the one of Gaussian model. The occurrence of the other models gradually decreased from the exponential to spherical, power, penta-spherical and logarithmic models. The latter is very rare fitted during the 30 yr. In reality, the logarithmic model is negative for small values of h, which rarely occurred or never in these catchments.

To avoid negative estimates in kriging, the model was changed to another one (Sect. 2.1.2). Using these seven
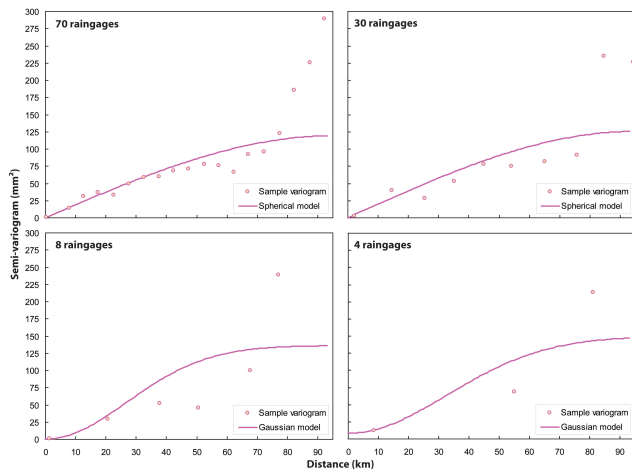
Fig. 4. Sample and theoretical semi-variograms for a day of winter (20 December 1993) made using 70, 30, 8 and 4 raingages.
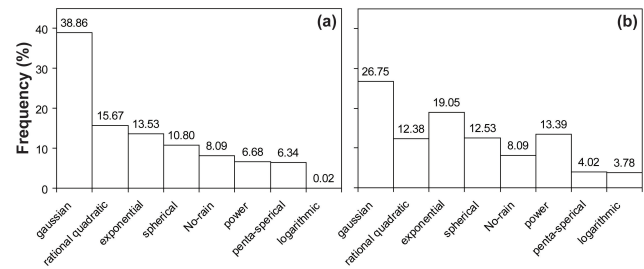
Table 1. Annual mean rainfall (in mm) of the Ourthe and Ambleve catchments, generated using six different methods and 70 raingages available in and surrounding the two catchments.

Methods	Average	Max	Min	At the highest point	At the lowest point
THI	1045	1421	840	1421	847
IDW	1045	1330	878	1300	881
ORK	1046	1334	862	1328	878
UNK	1052	1340	851	1340	881
KED	1065	1406	848	1406	848
OCK	1050	1327	847	1325	851

Fig. 5. Occurrence of variogram models used for kriging of the 30-yr dataset. **(a)**: the best fitted models; **(b)**: the chosen models avoiding negative kriging.

variogram models in this study, we avoided negative estimates for ordinary kriging even though using these seven different variogram models might not greatly improve the results, based on Fig. 2. After this operation, the frequency of Gaussian, rational quadratic and penta-spherical models decreased by 12.11 %, 3.29 % and 2.32 % respectively while the frequency of power, exponential, logarithmic and spherical models increased by 6.71 %, 5.52 %, 3.76 % and 1.73 % respectively (Fig. 5b). The change from one to another was significantly characterized by season. The frequency of Gaussian models decreased while the frequency of logarithmic model increased during spring and summer than autumn and winter (Fig. 6).

3.2 Spatial distribution of rainfall

The most straightforward method used was the Thiessen polygon (THI), whereby the value of the closest observation was simply assigned to each grid. The Thiessen polygon map showed the characteristics of the polygonal zones of influ-

ence around each raingages (Fig. 7). This method obviously provided an unrealistic discontinuous rain field at the border of each polygon, and did not show the true spatial variation of rainfall. The annual mean rainfall varied from 840 mm to 1421 mm at the highest elevation with a mean over the area of 1035 mm (Table 1).

For the Inverse Distance Weighting (IDW) map, the closest measured values had the most influence. IDW used a simple algorithm based on distance. In this study, we used inverse square distances to obtain the values at all grids in the catchment area. The map showed a distribution in more or less individual areas. Within these areas, there was usually a duck-egg shape corresponding to a high or low rainfall value (Fig. 7). The annual mean rainfall varied from 878 mm to 1330 mm with a mean value over area of 1045 mm (Table 1).

Instead of distance, kriging formed weights from surrounding measured values to predict values at each grid. However, the kriging weights for the surrounding measured points were more sophisticated than those produced by IDW. The kriging weights came from a semi-variogram that was developed by looking at the spatial structure of the data. The predictions of each grid were made based on the semi-variogram and the spatial arrangement of measured values that were nearby. The map generated by Ordinary Kriging (ORK) showed a relatively similar pattern to map obtained by IDW. However, the ORK map was smooth, presenting fewer duck-egg shapes (Fig. 7). The annual mean rainfall varied from 862 mm to 1334 mm, with an areal mean of 1046 mm (Table 1).

The trend modelled as a polynomial function of geographical coordinates was used to enhance the estimation by Universal Kriging (UNK) method. The influence of coordinates could make the map smoother than the map obtained by ORK (Fig. 7). However, the average areal rainfall was 1052 mm, a bit higher than the value obtained using previous methods. The annual mean rainfall varied from 851 mm to 1340 mm at the highest elevation (Table 1).
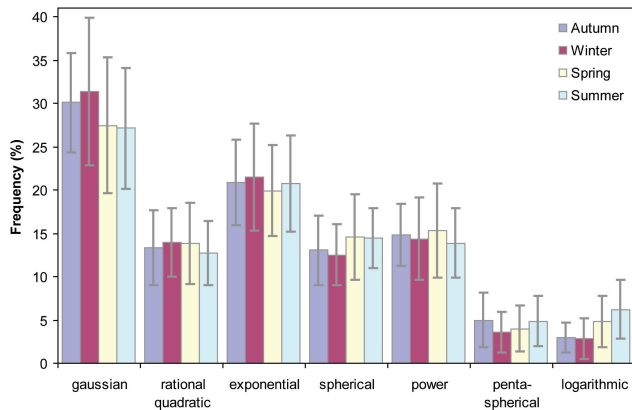
Fig. 6. Occurrence of the different models chosen for each season: bars represent the average percentage of the total number of rainy days and error bars are minus and plus the standard deviation of the 30-yr daily dataset.

Fig. 7. Annual mean rainfall of the Ourthe and Ambleve catchments, generated using six different algorithms and 70 raingages available in and surrounding the two catchments.

We also integrated the elevation information extracted from Digital Elevation Model (DEM) in order to improve the previous estimation by using the Kriging with an External Drift (KED) and Ordinary Cokriging (OCK). The KED used the elevations as the secondary variable to derive the local mean of rainfall (primary variable) while OCK took advantage of the correlation between the two variables (elevation and rainfall). Over the 30 yr (1976–2005), there were a total of 10063 rain days, on which the Pearson's coefficient was computed from the correlation between rainfall amount and elevation (Fig. 8). Of these 10063 rain days, 2087 rain days (20.74 %) had a Pearson's correlation coefficient higher than 0.5 and 181 rain days (1.8 %) had a Pearson's correlation coefficient lower than −0.5. In this study, a day was designated a "no-rain day" if rainfall for that day was equal to zero for all raingages of available data in the area, otherwise, the day was designated as a "rain day".

For KED, a visible high influence of topography over the estimated precipitation values appeared clearly in the map (Fig. 7). Overall, KED tended to overestimate the annual mean rainfall over the area (1065 mm). The annual mean rainfall varied from 848 mm at the lowest elevation to 1406 mm at the highest elevation (Table 1).

For the multivariate extension of kriging, Ordinary Cokriging (OCK) was used by incorporating the elevation derived from DEM as secondary information. The elevation was known in each grid and varied smoothly across the study area (Fig. 3). The map derived using the OCK technique exhibited nearly similar pattern to those derived using the UNK and ORK methods (Fig. 7). Moreover, the mean rainfall over the area was also in the same range as the mean of THI, IDW, ORK and UNK methods. The annual mean rainfall varied from 847 mm to 1327 mm with mean over area of 1050 mm (Table 1).
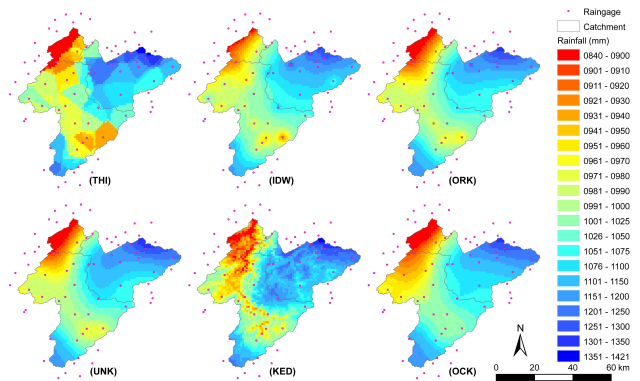
However, when the fewest gages were used for interpolation, UNK and KED provided very poor results (Fig. 9) and over-estimated the annual mean rainfall. The annual mean rainfall generated by UNK and KED varied from 558 mm and 298 mm to 2391 mm and 3719 mm, with a mean of 1141 mm and 1214 mm respectively.

3.3 Performance of daily rainfall interpolators

The number of gages was a factor for slightly reducing the RMSE (Fig. 10). There were little differences between geostatistical and IDW methods. However, the RMSE of estimates made using Thiessen polygon was clearly higher than those resulting from geostatistics and IDW. When the number of gages becomes very small, the RMSE becomes very high and the difference in the RMSE between the methods becomes larger. The Thiessen polygon provided the largest average RMSE value (2.81 mm), while ORK gave the lowest average RMSE value (2.42 mm). Other methods gave a somewhat higher average RMSE than ORK–IDW (2.44 mm), UNK (2.49 mm), KED (2.50 mm) and OCK (2.53 mm). But their RMSE values stayed much lower than those of the Thiessen Polygon.

For ORK, estimates based on more raingages tended to produce lower RMSE values. ORK outperformed the technique in most cases, while IDW provided lowest RMSE values where 8 and 20 gages were used. UNK provided the second lowest RMSE values where 50 and 60 raingages were used (Fig. 10).

However, only seven raingages were used for this performance evaluation. In order to validate the performance of these different methods, cross validations were performed for two period of very distinct rain pattern. For the daily rainfall with 70 and 30 raingages, Geostatistical and IDW methods still performed best out of the Thiessen algorithm considered in terms of RMSE, although small differences were found in some days particularly in summer (Fig. 11). However, for
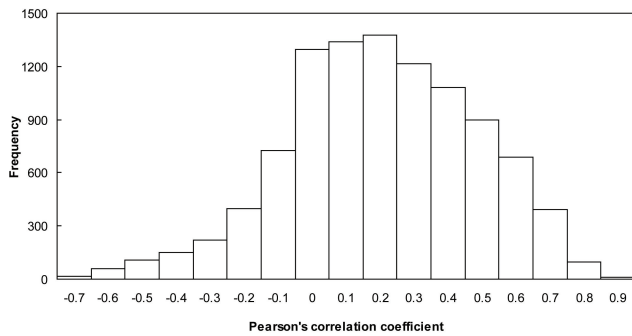
Fig. 8. Histogram of Pearson's correlation coefficients computed from the correlation between daily rainfall (1976–2005) and altitude extracted from DEM.

Fig. 9. Annual mean rainfall of the Ourthe and Ambleve catchments, generated using six different algorithms and 4 raingages.

the interpolation with 4 raingages, UNK and KED gave very high RMSE not only for the cross validation (Fig. 11) but also for the seven validation gages (Fig. 10).

For all methods, the high RMSE are found mostly at high elevation stations and the stations located far from centre of the catchments (Fig. 11). However, the Thiessen polygon always presented higher RMSE over other methods. As suggested by the data taken on one summer day (14 July 1994), the rain pattern is convective, and the high rain amounts were present at only small part of the catchments. High error of Thiessen polygon were found for the stations in area where it rained and no error was found at the no-rain stations (Fig. 12). But for the IDW and geostatistical methods, the errors were found at nearly all raingages. However, the cross validation RMSE of the IDW and geostatistical methods, at this day, appeared clearly lower than the error of the Thiessen polygon (Fig. 11).

4 Discussion

Results from the application of different algorithms provided some insights in terms of strengths and weaknesses, and in terms of the applicability of the deterministic and geostatistical methods to daily rainfall made using different densities of raingages in the Ourthe and Ambleve catchments. All the algorithms were able to produce 30-yr daily rainfall on the catchment grids for a distributed hydrological model, ranging from the most straightforward (THI) to the most complex method (OCK).

Negative kriging weights applied to extreme values can result in kriging estimates outside the range of observed data. In many situations, negative kriging lead to nonphysical estimates (Deutsch, 1996). This feature may be necessary for daily rainfall. To avoid negative estimates in kriging in the present study, the variogram model changed from one another. The frequency change of models varied according to the seasonal pattern of rainfall. A significant difference could be found for Gaussian and logarithmic models (Fig. 6).
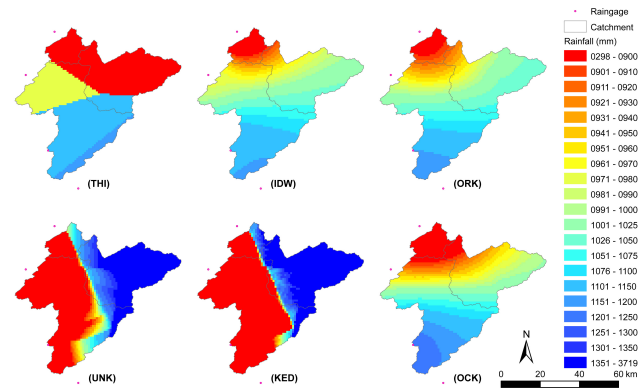
The logarithmic model frequency was higher while Gaussian model frequency was lower during spring and summer than autumn and winter. This means that the negative estimates in kriging occurred more during summer than winter. In the Ourthe and Ambleve catchments, the more convective type of rain appeared mostly during summer when high data values were locally present. It is an outlying data event which can lead to negative estimates in kriging (Deutsch, 1996).

For both validation approaches, the geostatistical and IDW methods outperform the Thiessen polygon method for the Ourthe and Ambleve catchments. Velasco-Forero et al. (2009) recommended analysing the sensitive of their different interpolation technique to the density of raingage network. Current study has focused on this recommendation even if with different method of variogram modelling. It is found that raingage density was one of the factors in determining the performance of such interpolation method (Fig. 10). The use of large number of raingages did not improve result. Too few stations produced poor interpolation results. The optimal number of the stations was between 8 and 70. Here, the best methods were ORK and IDW for daily rainfall from 1976 to 2005 in the Ourthe and Ambleve catchments. The present study also shows that IDW had a smaller error of estimates than ORK and UNK when using 30, 20, 8 and 4 gages (Fig. 10). IDW weight is the inverse distance of the neighbour points while kriging weight is determined by semi-variogam using spatial relationship of both distances and values of the neighbour points. So the IDW is the most adequate because the stations used for interpolation of these cases might be close to the seven validation stations.

There was a sudden decline in Thiessen-polygon RMSE value when using eight raingages for interpolation (Fig. 10) because the eight raingages are close to the seven raingages selected for validation. Although, RMSE values, provided by geostatistical and IDW methods, gradually decreased according to the number of raingages used for the interpolation (Fig. 10). In general, the values of points close to the sample points were more likely to be similar than those that
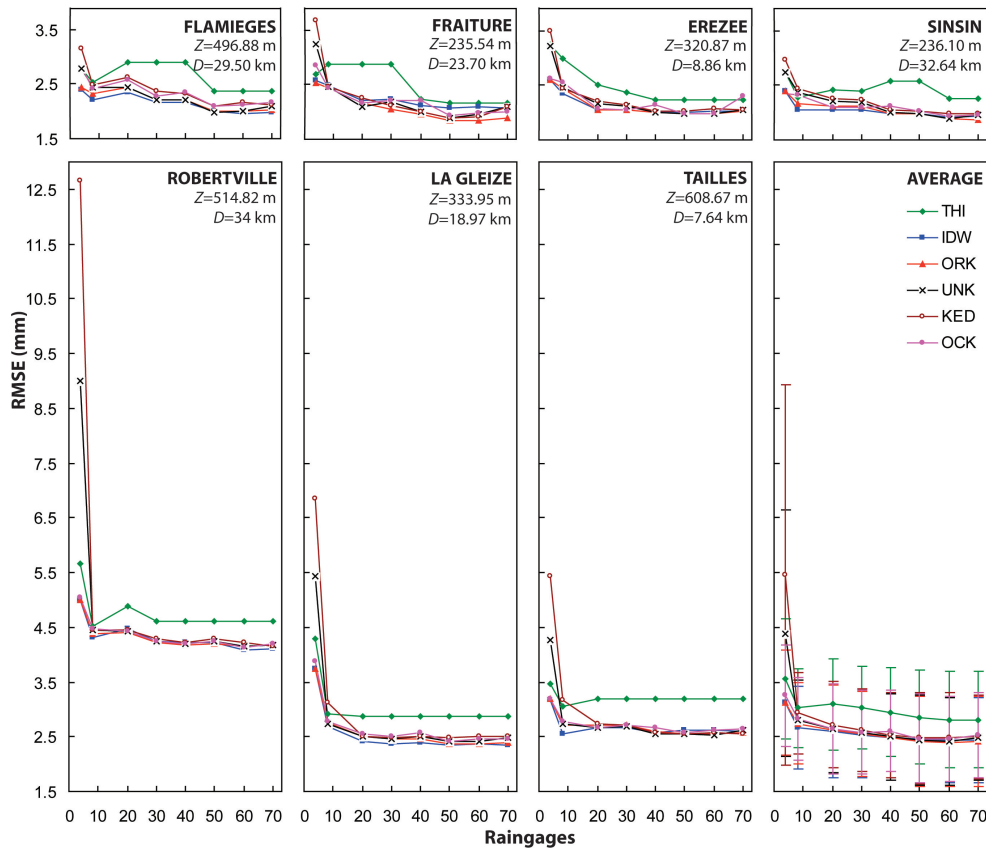
Fig. 10. Evolution of RMSE values of different methods and validation gages according to number of gages used for the spatial interpolation: the points (low-right) are the mean and error bars are plus and minus standard deviation of the seven validation raingages; Z is the elevation of the raingages; D is the distance from the raingages to the centre of the two catchments.

are further apart. The Thiessen polygon ignored the pattern of spatial dependence and considered only one measurement, whereas IDW and geostatistics were respectively based on the surrounding measured values and statistical models that included autocorrelation – the statistical relationship between the measured points.

It is interesting to note that UNK and KED showed some limitations and tended to over-estimate the mean rainfall over the catchment area. In particular, the most critical case in this study used the fewest raingages, which were only at the low-elevation part of the catchment area. UNK and KED produced very poor results in terms of both rainfall distribution and accuracy from both validation raingages (Fig. 10) and cross validation (Fig. 11). The rainfall distributions were very poorly represented (Fig. 9). The RMSE values were very high. This can be explained by the extrapolation of the UNK and KED outside the range of data. The values of the local trend (coordinate function) and sampled secondary variable (elevation) were outside the range of the values at locations where the primary variable were also sampled. The maximum elevation of DEM was 693 m (Fig. 3) while the maximum raingage elevation was 552 m. The external drift

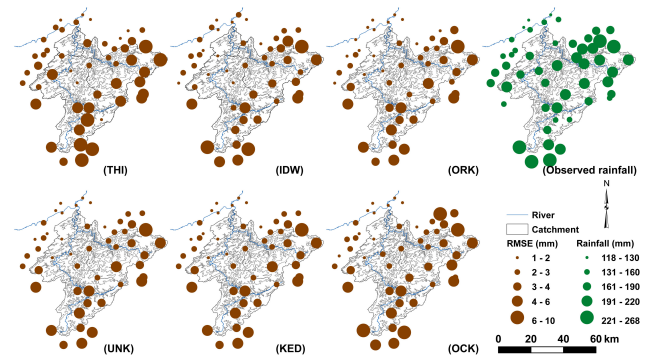parameters must have an adequate range to avoid extrapolation. The data scarcity was more of a limiting factor.

Integrating elevation into KED and OCK for spatial interpolation did not really lead to a smaller error of estimates here, it showed, instead, the highest RMSE value between IDW and other geostatistical methods for most of the gage-degenerated cases (Fig. 10). This was because of the poor correlation between elevation and daily rainfall (Fig. 8). Certainly, differences in the time step used for interpolation could contribute to the difference in the result of the present study and those of studies of Goovaert (2000) and Lloyd (2005), which used monthly and annual time steps. Accounting for elevation using multivariate geostatistical algorithms (KED and OCK) generally reduces the ORK prediction error as long as the correlation coefficient is larger than 0.75. The benefit of multivariate techniques can, therefore, become marginal if the correlation between rainfall and elevation is too small (Goovaert, 2000). When the observation time steps are less than one month, the location relationship between precipitation and altitude is likely to be less obvious (Lloyd, 2005). In our study, the rainfall accumulations were obtained in shorter time steps (daily) while daily
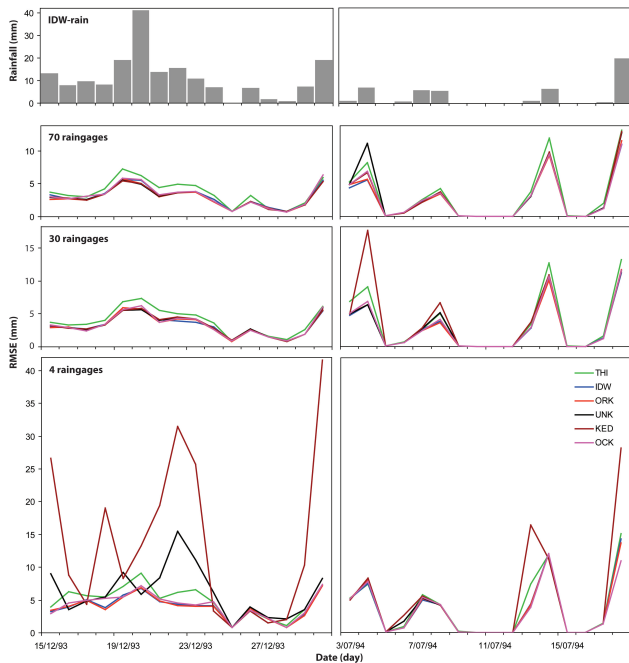
Fig. 11. Cross validation of different interpolation methods for a period of winter (left) and a period of summer (right). The first graph presents the areal rainfall interpolated by IDW; the second graph present the daily RMSE of the methods made using 70 raingages; the third graph present the daily RMSE of the methods made using 30 raingages; the fourth graph present the daily RMSE of the methods made using 4 raingages.

Fig. 12. RMSE and total observed rainfall maps for series of a selected period of winter (16 to 30 December 1993). The background is the contour line of the catchments.

rainfall data is routinely handled at weather observation stations and often an important meteorological input to water resources and agricultural modelling systems. Unfortunately, the correlations between daily rainfall and elevation were relatively small for most of the rain days (Fig. 8). Hence, ORK and IDW provided better results than UNK, KED and OCK. In previous studies regarding daily rainfall (Schuurmans et al., 2007), KED and OCK were more accurate than ORK, especially for larger extents with lower densities of raingages. But they used the added value of radar which generally is well correlated with rainfall from raingage, certainly thanks to the same type of variable.

Among the seven validation raingages, the Robertville raingage with a high elevation (514.82 m) and situated near the catchment's border provided systematically the highest RMSE value in all of the gage-degenerated cases (Fig. 10). Furthermore, Tailles, a raingage with the highest elevation (608.67 m) and situated in the boundary (peak line) between the two catchments provided the second highest RMSE value. For the cross validation, the high errors were present mostly at the high elevation part and at the boundary of the catchment (Fig. 11). This can be explained by the effect of high elevation and the positions of those raingages which are at the extremity of the zone. The least errors occur over flat

plains and the largest over mountainous area (Basistha et al., 2008). In general, the average rainfalls are found higher at high elevation than low elevation.

It is also important to stress that our geostatistical methods use seven variogram models to avoid negative rainfall. The results showed that Kriging using only spherical model provided a slightly better result (0.05 mm of RMSE difference average) than those using seven models for ORK, UNK and KED but not for OCK which gained 0.12 mm of RMSE by using seven variogram models (Fig. 13). This confirms that the use of these other approaches did not really improve much the results.

Regarding the interpolation evaluations, raingage values data are normally used for the evaluation of interpolated rainfalls. In this study, data from certain raingages were not included in the raw data set, but they used for evaluation purposes. The excluded evaluation data were randomly taken at a wide range of elevation and spread over the catchment, and were not used in the interpolation process. In order to increase the number of evaluation points we used cross validation techniques for short period. This is definitely a weakness of the methodology. However, the results obtained using the cross validation show the same trend in RMSE as evaluation of the seven raingages, confirming that the use of the modified selection criteria did not affect the results significantly. The extended validation was justified for statistical reasons.
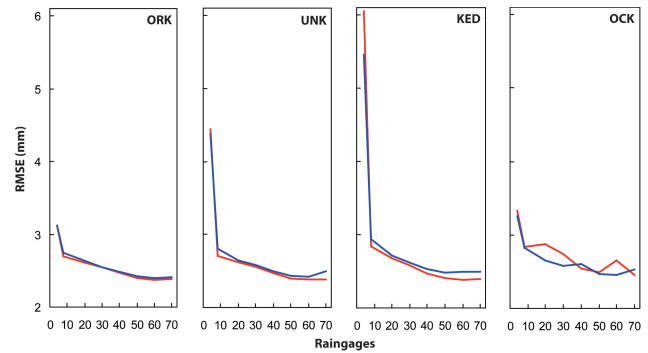
5 Conclusions

The geostatistical algorithms used a daily based variogram model chosen from the seven variogram models to avoid negative rainfall results. The multivariate geostatistics could incorporate the elevation as secondary variable. All the algorithms were able to produce successfully the long series of daily rainfall on all the 1 km^2 grid of the catchment area according to different densities of raingages, ranging from the
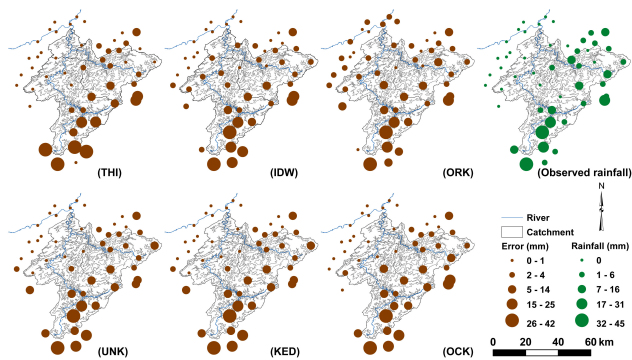
Fig. 13. Absolute error and observed rainfall maps for a day of summer (14 July 1994). The background is the contour line of the catchments.

Fig. 14. Comparison between Kriging using seven variogram models (blue line) and Kriging using only spherical variogram model and simply setting negative rainfall to zero (red line).

most available raingage to very sparse raingage. The main results and conclusions can be summarised as follows:

– Between the seven variogram models used, the Gaussian model was the most frequently best fitted, which should be recommended for the spatial interpolation of daily rainfall if only one model would be applied.

– Using seven variogram models could avoid negative daily rainfall in ordinary kriging. It is found that the negative estimates of kriging occurred for convective more than stratiform rain.

– The performance of the different methods varied slightly according to the density of raingages, particularly between 8 and 70 raingages but it was much different for interpolation using very sparse raingages.

– Geostatistical and IDW algorithms significantly outperformed simple techniques like the Thiessen polygon which is commonly used in various hydrological models. Here, the Thiessen polygon clearly provided a discontinuous rain field at the border of each polygon, and did not show the true spatial variation of rainfall. For both validation approaches, the RMSE values were always highest in all gage-degenerated cases.

– Care should be taken in using UNK and KED when interpolating with very few neighbourhood sample points. These methods can extrapolate outside the range of data values and cause a poor result. These recommendations complement the results on using UNK and KED for daily rainfall reported in the literature.

– Between the geostatistical and IDW methods, KED and OCK using elevation were not supposed to be the improved methods because the correlation of daily rainfall and elevation was small for most of the rain days. However they were not much different if the data was not very sparse.

– ORK was considered to be the best and most robust method since it provided lowest RMSE value for nearly all cases. This method was followed by IDW. However, the IDW method was much simpler than complex geostatistical methods which require a lot of computation time.

– ORK, UNK and KED using only spherical model provided a slightly better result but OCK using seven variogram models achieved better result.

These different techniques have been presented, in science, as a prerequisite to their uses in hydrological models. Such analyses in this study can be vital to scientists, engineers, hydrologists, and decision makers alike. A subject that remains to be explored is what methods that produce daily rainfall for a distributed hydrological model can provide the best results for stream flow simulation.

Edited by: R. Uijlenhoet

References

Andréassian, V., Perrain, C., Michel, C., Usart-Sanchez, I., and Lavabre, J.: Impact of imperfect rainfall knowledge on efficiency and the parameters of watershed models, J. Hydrol., 250, 206–223, 2001.

Angulo-Martínez, M., López-Vicente, M., Vicente-Serrano, S. M., and Beguería, S.: Mapping rainfall erosivity at a regional scale: a comparison of interpolation methods in the Ebro Basin (NE Spain), Hydrol. Earth Syst. Sci., 13, 1907–1920, doi:10.5194/hess-13-1907-2009, 2009.

Angulo-Martínez, M. and Beguería, S.: Estimating rainfall erosivity from daily precipitation records: A comparison among methods using data from the Ebro Basin (NE Spain), J. Hydrol., 379, 111–121, 2009.

Aronica, G. and Ferro, V.: Rainfall erosivity over the Calabrian region/Erosivité des précipitations en Calabre, Hydrolog. Sci. J., 42(1), 35–48, 1997.

Attorre, F., Alfo, M., Sanctis, M. D., Francesconi, F., and Bruno, F.: Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale, Int. J. Climatol., 27, 1825–1843, 2007.

Basistha, A., Arya, D. S., and Goel, N. K.: Spatial Distribution of Rainfall in Indian Himalayas – A case study of Uttarakhand Region, Water Resour. Manag., 22, 1325–1346, 2008.

Beek, E. G.: Spatial variability and interpolation of daily precipitation amount, Stochastic Hydrol. Hydraul., 6, 304–320, 1992.

Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299, 166–179, 2004.

Boer, E. P. J., de Beurs, K. M., and Hartkamp, A. D.: Kriging and thin plate splines for mapping climate Variables, Int. J. Appl. Earth Obs. Geoinf., 3, 146–154, 2001.

Borga M. and Vizzaccaro, A.: On the interpolation of hydrologic variables: Formal equivalence of multiquadratic surface fitting and kriging, J. Hydrol., 195, 160–171, 1997.

Buytaert, W., Celleri, R., Willems, P., Bièvre, D. B., and Wyseure, G.: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes, J. Hydrol. 329, 413–421, 2006.

Campling, P., Gobin, A., and Feyen, J.: Temporal and spatial rainfall analysis across a humid tropical catchment, Hydrol. Process. 15, 359–375, 2001.

Carrera-Hernández, J. J. and Gaskin, S. J.: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico, J. Hydrol., 336, 231–249, 2007.

Caruso, C. and Quarta, F.: Interpolation Methods Comparison, Comput. Math. Applic., 35, 109–126, 1998.

Chaubey, I., Haan, C. T., Grunwald, S., and Salisbury, J. M.: Uncertainty in the model parameters due to spatial variability of rainfall, J. Hydrol., 220, 48–61, 1999.

Chilès, J. P. and Delfiner, P.: Geostatistics: Modelling Spatial Uncertainty, John Wiley & Sons, New York, 1999.

Chow, V. T.: Handbook of applied hydrology: A compendium of water-resources technology, McGraw-Hill, Inc. 1964.

Cole, S. J. and Moore, R. J.: Hydrological modelling using raingage and radar-based estimators of areal rainfall, J. Hydrol., 358, 159–181, 2008.

Collischonn, B., Collischonn, W., and Tucci, C. E. M.: Daily hydrological modeling in the Amazon basin using TRMM rainfall estimates, J. Hydrol., 360, 207–216, 2008.

Cressie, N.: Fitting Variogram Models by Weighted Least Squares, Math. Geol. 17, 563–586, 1985.

Cressie, N.: Statistics for spatial data, Wiley, New York, 900 pp., 1991.

Demarcin, P., Degre, A., Smoos, A., and Dautrebande, S.: Projet ERRUISSOL. Cartographie numérique des zones à risque de ruissellement et d'érosion des sols en Région wallonne (Belgique), Rapport final de convention DGO3, Unité d'hydrologie et hydraulique agricole, Faculté Universitaire des Sciences Agronomiques de Gembloux, 55 pp., 2009.

Deutsch, C. V.: Correcting for negative weights in ordinary kriging, Comput Geosci, 22, 765–773, 1996.

Diggle, P. J. and Ribeiro Jr., P. J.: Model-based Geostatistics, Springer Series in Statistics, Springer Science, Business Media, LLC, 228 pp., 2007.

Dirks, K. N., Hay, J. E., Stow, C. D., and Harris, D.: High-resolution of rainfall on Norfolk Island, Part II: Interpolation of rainfall data, J. Hydrol. 208, 187–193, 1998.

Faurès, J. M., Goodrich, D. C., Woolhiser, D. A, and Sorooshian, S.: Impact of small-scale spatial variability on runoff modeling, J. Hydrol., 173, 309–326, 1995.

Gabellani, S., Boni, G., Ferraris, L., Hardenberg, J. V., and Provenzale, A.: Propagation of uncertainty from rainfall to runoff: A case study with a stochastic rainfall generator, Adv. Water Resour., 30, 2061–2071, 2007.

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Oxford University Press, New York, 483 pp., 1997.

Goovaerts, P.: Geostatistical tools for characterizing the spatial variability of microbiological and physico-chemical soil properties, Biol. Fert. Soils., 27, 315–334, 1998.

Goovaerts, P.: Using elevation to aid the geostatistical mapping of rainfall erosivity, Catena, 34, 227–242, 1999.

Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, J. Hydrol., 228, 113–129, 2000.

Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, J. Hydrol. 332, 144–157, 2007.

Hevesi, J., Istok, J., and Flint, A.: Precipitation Estimation in Mountainous Terrain Using Multivariate Geostatistics. Part I: Structural Analysis, J. Appl. Meteorol. 31, 661–676, 1992.

Hoyos, N., Waylena, P. R., and Jaramillo, A.: Seasonal and spatial patterns of erosivity in a tropical watershed of the Colombian Andes, J. Hydrol., 314, 177–191, 2005.

Isaaks, E. H. and Srivastava, R. M.: An Introduction to Applied Geostatistics, Oxford University Press, New York, 561 pp., 1989.

Johnson, G. L. and Hanson, C. L.: Topographic and atmospheric influences on precipitation variability over a mountains watershed, J. Appl. Meteor., 34, 68–87, 1995.

Kobold, M. and Sušelj, K.: Precipitation forecasts and their uncertainty as input into hydrological models, Hydrol. Earth Syst. Sci., 9, 322–332, doi:10.5194/hess-9-322-2005, 2005.

Lanza, L. G., Ramírez, J. A., and Todini, E.: Stochastic rainfall interpolation and downscaling, Hydrol. Earth Syst. Sci., 5, 139–143, doi:10.5194/hess-5-139-2001, 2001.

Leander, R., Buishand, T. A., van den Hurk, B. J. J. M, and de Wit, M. J. M: Estimated changes in flood quantiles of the river Meuse from resampling of regional climate model output, J. Hydrol.,

351, 331–343, 2008.

Lloyd, C. D.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, J. Hydrol., 308, 128–150, 2005.

Lu, G. Y. and Wong, D. W.: An adaptive inverse-distance weighting spatial interpolation technique, Comput. Geosci., 34, 1044–1055, 2008.

Matheron, G.: The Theory of Regionalised Variables and its Applications: Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau N°5, The Ecole Nationale Supérieure des Mines de Paris, Paris, France, 1971.

Marquínez, J., Lastra, J., and García, P.: Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis, J. Hydrol. 270, 1–11, 2003.

Moulin, L., Gaume, E., and Obled, C.: Uncertainties on mean areal precipitation: assessment and impact on streamflow simulations, Hydrol. Earth Syst. Sci., 13, 99–114, doi:10.5194/hess-13-99-2009, 2009.

Nalder, I. A. and Wein, R. W.: Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest, Agric. For. Meteorol. 92, 211–225, 1998.

Nyssen, J., Vandenreyken, H., Poesen, J., Moeyersons, J., and Deckersd, J.: Rainfall erosivity and variability in the Northern Ethiopian Highlands, J. Hydrol., 311, 172–187, 2005.

Pardo-Igúzquiza, E.: MLREML4: A Program for the inference of the power variogram model by maximum likelihood and restricted maximum likelihood, Comput. Geosci., 24, 537–543, 1998.

Ruelland, D., Ardoin-Bardin, S., Billen, G., and Servat, E.: Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa, J. Hydrol., 361, 96–117, 2008.

Schiemann, R., Erdin, R., Willi, M., Frei, C., Berenguer, M., and Sempere-Torres, D.: Geostatistical radar-raingauge combination with nonparametric correlograms: methodological considerations and application in Switzerland, Hydrol. Earth Syst. Sci., 15, 1515–1536, doi:10.5194/hess-15-1515-2011, 2011.

Schuurmans, J. M. and Bierkens, M. F. P.: Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, Hydrol. Earth Syst. Sci., 11, 677–693, doi:10.5194/hess-11-677-2007, 2007.

Schuurmans, J. M., Bierkens, M. F. P., Pebesma, E. J., and Uijlenhoet, R.: Automatic prediction of high-resolution daily rainfall fields for multiple extents: The potential of operational radar, J. Hydrometeorol., 8, 1204–1224, 2007.

Singh, V. P.: Effect of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, Hydrol. Process., 11, 1649–1669, 1997.

Spadavecchia, L. and Williams, M.: Can spatio-temporal geostatistical methods improve high resolution regionalisation of meteorological variables?, Agr. Forest Meteorol., 149, 1105–1117, 2009.

Syed, K. H., Goodrich, D. C., Myers, D. E., and Sorooshian, S.: Spatial characteristics of thunderstorm rainfall fields and their relation to runoff, J. Hydrol., 271, 1–21, 2003.

Teegavarapu, R. and Chandramouli, V.: Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records, J. Hydrol., 312, 191–206, 2005.

Todini, E., Pellegrini, F., and Mazzetti, C.: Influence of parameter estimation uncertainty in Kriging: Part 2 – Test and case study applications, Hydrol. Earth Syst. Sci., 5, 225–232, doi:10.5194/hess-5-225-2001, 2001.

van de Beek, C. Z., Leijnse, H., Torfs, P. J. J. F., and Uijlenhoet, R.: Climatology of daily rainfall semi-variance in The Netherlands, Hydrol. Earth Syst. Sci., 15, 171–183, doi:10.5194/hess-15-171-2011, 2011.

Velasco-Forero C. A., Sempere-Torres, D., Cassiraga E. F., and Gómez-Hernández, J. J.: A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data, Adv. Water Resour., 32, 986–1002, 2009.

Verworn, A. and Haberlandt, U.: Spatial interpolation of hourly rainfall effect of additional information, variogram inference and storm properties, Hydrol. Earth Syst. Sci., 15, 569–584, doi:10.5194/hess-15-569-2011, 2011.

Vicent-Serrano, S. M., Saz-Sanchez, M. A., and Cuadrat, J. M.: Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature, Clim. Res., 24, 161–180, 2003.

Webster, R. and Oliver, M. A.: Geostatistics for Environmental Scientists, Statistics in Practice Series, John Wiley & Son Ltd., 315 pp., 2007.