Conditional simulation schemes of rain fields and their application to rainfall–runoff modeling studies in the Sahel

Théo Vischel a,*, Thierry Lebel a, Sylvain Massuel b,1, Bernard Cappelaere b

a Laboratoire d’Etude des Transferts en Hydrologie et Environnement (UMR 5564), IRD, Grenoble, France
b Laboratoire d’étude des Transferts en Hydrologie et Environnement (UMR 5564), IRD, Grenoble, France

Available online xxxx

ARTICLE INFO

Keywords:
Stochastic simulation of rain fields
Conditional simulations
Geostatistics
Rainfall–runoff modeling
Sahel

SUMMARY

In regions characterized by a great inter-annual variability or by decadal-scale changes of the rainfall regime, the simulation of long series of rainfall events is an efficient way to explore the runoff fluctuations or modifications resulting from this rainfall variability. In a context of great uncertainty regarding the Sahelian rainfall regime in a changing climate, a coherent stochastic framework is presented here to produce high spatial resolution rain fields in order to force a rainfall–runoff model and to perform sensitivity analyses. The focus of the paper is on the comparison of various conditioning methods reflecting the various types of information available for the study of past situations (data from rain gage and satellite) as well as of future scenarios (outputs of atmospheric models).

Various types of rainfall simulations are performed over a 13 year period, using four levels of conditioning information obtained from a 15 gauge network covering a 60 km² region. These simulations are then used as inputs to a Hortonian rainfall–runoff model. The simulation relevance is first assessed by studying the simulated rain field series (event time-step mean characteristics, seasonal cycle and inter-annual variability) in comparison with reference rain fields estimated by kriging. This shows that the conditioning of the simulations, even by a minimal information provided by a unique station, is of great relevance for constraining the stochastic dispersion and thus to retrieve the rainfall variability at the considered scales. Significant differences are reported between runoff obtained by the different types of created rain fields, one of the most noticeable being that runoff obtained from kriging is 25% lower than runoff obtained from point conditional simulations. The results confirm the sensitivity of Hortonian hydrological systems to rainfall intensity and particularly point out the importance of representing realistic spatial rainfall patterns to force hydrological models.

Introduction

One key motivation for implementing a long term hydrometeorological monitoring system in West Africa is to obtain high space-time resolution data in order to calibrate and validate the models used to understand and quantify the hydrological impact of the variability and/or long term change of the rainfall regime. The West African drought that started at the end of the 1960s had its most dramatic effects in the Sahel, peaking in the mid 1980s. As shown in Lebel and Ali (2009) a rainfall deficit is still observed today over the Sahel, even though in a somewhat abated way, producing a streamflow deficit that is twice as large over the large region of semi-arid climate – for at least four reasons: (i) rainfall is of convective origin and thus highly variable in space and time; that the link between rainfall variability and runoff/streamflow variability is scale dependent, which is in line with the results of several studies (see e.g. Goodrich et al., 1997; Sivapalan et al., 2002; Vivoni et al., 2007). This underlines the need to use, when possible, hydrological models representing explicitly the hydrological processes at the fine scales at which they occur and interact (e.g. Abbott et al., 1986; Refsgaard and Storm, 1995; Peugeot et al., 2003).

One of the most important and constant issues in the use of these hydrological models is to use rainfall input data at scales compatible with the scales required by the processes being modeled. Numerous studies report the risk of biasing the modeling of the hydrological systems when using rainfall inputs at inappropriate spatial scales (Michaud and Sorooshian, 1994; Faurès et al., 1995; Finnerty et al., 1997; Winchell et al., 1998; Liang et al., 2004; Vischel and Lebel, 2007; Segond et al., 2007). Potential scale mismatch problems are of particular concern in the Sahel – a region of semi-arid climate – for at least four reasons: (i) rainfall is

* Corresponding author. Tel.: +33 476 827 089; fax: +33 476 825 286.
E-mail address: theo.vischel@hmg.inpg.fr (T. Vischel).

1 Present address: CSIRO L&W, Private bag 5, Wembley 6913, WA, Australia.
(ii) runoff is produced through Hortonian mechanisms, which means that rainfall intensities within the event are as important as the total event rainfall; (iii) the spatial distribution of these intensities is another key factor, especially when considering that the density of the national rain gage networks has continuously degraded over the past 15 years (Ali et al., 2005); (iv) rainfall estimates from satellites or outputs of atmospheric models are either too inaccurate or available at a too coarse resolution to be directly used to force fine scale hydrological models.

As a result, the ability to simulate fine resolution rain fields remains an important need for studying the hydrological impact of rainfall regime changes affecting the patterns of spatial and time rain distributions, in particular in the framework of climate change research. Given that climate scenarios are especially dispersed for the Sahel (IPCC Report, 2007), the aim of the work presented here is to provide a unified framework for simulating the impact on runoff of various possible rainfall regime modifications. The focus is on testing the impact of different conditioning information on the simulation process, the paper being organised as follows:

1. Four different simulation schemes of event rain fields are implemented and compared in terms of their ability to reproduce observed event rainfall characteristics; these schemes differ by the way they are conditioned to point observations or to areal values, corresponding to various situations in term of data availability (for instance a single long term rainfall series may be the only source of information we have over several decades in the past; or we might want to produce fine resolution rain fields from the coarse resolution output of an atmospheric model simulating the rainfall regime of the future).
2. The simulations are conditioned by the observations of 545 rain events spanning 13 years thus allowing for comparison of the mean event characteristics, mean seasonal cycle and inter-annual variability of the rain fields series produced by each simulation scheme.
3. The simulated event rain fields are used as input to a Hortonian rainfall–runoff model in order to study the sensitivity of the runoff output to the rainfall forcing used.

Rainfall simulation schemes

A brief review of possible simulation schemes

A variety of stochastic rainfall models have been developed to simulate the spatial variability of rainfall. These models try to meet various requirements that are not easy to conciliate such as: making use of observations that are not necessarily covering the whole spectrum of scales involved in rainfall variability, representing the features of rain variability that are essential according to the end use of the hydrological model applications, allowing for a robust estimation of the parameters, conditioning the simulations by any information external to the simulation process. Rain cell models (e.g. Rodríguez-Iturbe and Eagleson, 1987; Kavvas et al., 1987; Féral et al., 2006) are an example of models too complex for the case treated here: a robust estimation of their large number of parameters is impossible with the recording rain gage data available to us (another impediment of these models is the complexity of conditioning by both point and areal values). Scale-invariance models (e.g. Schertzer and Lovejoy, 1987; Over and Gupta, 1994; Molnar and Burlando, 2005) are an attractive alternative since, with a small number of parameters they allow a straightforward conditioning by areal values; however these models require an accurate description of rainfall variability over a range of scales which was not obtainable with our rain gage network. In our context, the family of point rainfall models coupled to the simulation of a spatial correlation structure appears to provide a good compromise between robustness and complexity. In this family, the multi-site rainfall models are promising (e.g. Wilks, 1998; Cowpertwait et al., 2002; Mehrotra et al., 2006) but they are still numerically too expensive for simulating a large number of rain events on fine resolution grids at sub-daily time-steps. This is not the case for the meta-Gaussian models (e.g. Bell, 1987; Shah et al., 1996; Bouvier et al., 2003) that are based on the simulation of Gaussian random fields transformed by statistical functions into non-Gaussian rain fields. They are relatively simple to implement and make use of a limited number of parameters that can be inferred from point measurements. Moreover, a theoretical framework exists to condition the simulations by both point and areal values.

The model used in our study was initially developed by Lebel et al. (1998) to represent the structure of the ground rain fields associated with the Sahelian Organized Convective Systems (OCSs) that explain about 90% of the annual rainfall in the region. It is a space-time model involving three steps: (1) the simulation of the event rain field (meaning the spatially distributed rainfall depth accumulated over the whole event duration) based on meta-Gaussian random field simulations, (2) the definition of the dynamics of the simulated OCS by simulating at each grid point a time of arrival of the system, (3) the temporal disaggregation of the event rainfall depth according to a synthetic hyetograph representative of the intra-event variability of rainfall intensities. The main theoretical features of step 1 are dealt with in Guillot (1999), while some improvements of the temporal disaggregation (step 3) were proposed by Balme et al. (2006).

Non-conditional simulation of rain fields in the meta-Gaussian framework

The simulated variable is assumed to be a random variable entirely defined by its spatial structure (spatial covariance function or variogram) and its marginal cumulative distribution function (cdf – describing the frequency distribution of the values at point locations). The simulation of random fields characterized by any spatial structure function is relatively simple as long as the distribution of the simulated variable is Gaussian (Guillot, 1999). The Turning Bands Method (Mantoglou and Wilson, 1982), the Matrix Decomposition Method (Alabert, 1987) or the Direct Fourier Transform (Gutjahr, 1989) are three of the most classical methods used for generating Gaussian random fields. The generation of non-Gaussian random fields is less straightforward but a method was proposed by Journel and Huijbregts (1978) consisting of the two following steps:

1. The transformation of the Gaussian field $Y$ into a non-Gaussian field $Z$ by using an anamorphosis function $\phi$, defined so that the non-Gaussian field can be written as:

$$Z(x) = \phi(Y(x))$$  \hspace{1cm} (1)

2. An illustration of the use of the anamorphosis function is given in Fig. 1. According to Guillot (1999), an anamorphosis function $\phi$ can be derived from any cdf. However as $\phi$ may also transform the spatial structure function $S_{y}$ of the Gaussian field, one must properly define $S_{y}$ in order to obtain the expected spatial structure $S_{z}$ after anamorphosis (see details in Appendix A).
ues can be deduced from Eq. (2) as follows: the cdf function thus the corresponding Gaussian conditioning values can be assigned inside the interval $[0, S_1]$. Every element of the vector of non-zero values and the complementary vector of zero values. Every element of the vector of zero values.

Simulation of non-Gaussian random fields conditioned by point values

Again, the conditioning of random fields by point values is relatively easy to implement if the simulated variables are Gaussian (Lantuejoul, 1994). A few complex methods exist to simulate non-Gaussian random fields conditioned by point values (Emery, 2002). The simpler anamorphosis procedure is made of the following steps: (i) transforming the conditioning non-Gaussian values into Gaussian values, (ii) using the transformed values to condition the Gaussian random fields, (iii) transforming the conditioned Gaussian random fields into non-Gaussian random fields, based on Eq. (1). The first two points are detailed hereafter.

Transformation of non-Gaussian conditioning values into Gaussian values

Case of continuous non-Gaussian cdf. If the cdf of the simulated variable is continuous, the transformation of the non-Gaussian conditioning values into Gaussian values can be achieved by using the inverse of the anamorphosis function $\phi^{-1}$ (defined in Eq. (1)). Then:

$$Y(x_p) = \phi^{-1}(F(x_p))$$

where $x_p = (x_1, x_2, ..., x_n)$ is the vector of the location of the n-conditioning points.

Case of discontinuous non-Gaussian cdf. In the case of a discontinuous distribution of the simulated variables, the anamorphosis function $\phi$ is no longer injective. Eq. (2) is only valid on the continuous interval of the cdf. For the transformation of the atom values, several Gaussian values can be potentially assigned. An illustration of this problem is given in Fig. 2 for the case of a distribution with an atom at zero, which is the case for the application to Sahelian rain fields presented here.

The vector of conditioning points $Z(x_p)$ is separated into $Z(x_p^m)$ a vector of non-zero values and the complementary vector $Z(x_p^n)$ of zero values. Every element of $Z(x_p^m)$ is in the continuous part of the cdf function thus the corresponding Gaussian conditioning values can be deduced from Eq. (2) as follows:

$$Y(x_p) = \phi^{-1}(Z(x_p^m))$$

Three methods are then proposed for estimating the vector $Y(x_p^m)$:

(I) The first method (called hereafter $m_1$) consists of assigning to all elements of the vector $Y(x_p^m)$ the unique value $S_1$ defined by $S_1 = N^{-1}(1/F_0)$ where $N$ is the Gaussian distribution function and $F_0$ is the frequency of zero values.

(II) The second method ($m_2$) consists of randomly simulating the elements of the vector $Y(x_p^m)$ inside the interval $[-\infty, S_1]$.

(III) The third method ($m_3$) is more complex but is the most rigorous. It is based on the Gibbs sampling algorithm (Geman and Geman, 1984). The vector $Y(x_p^m)$ is estimated by iteratively kriging the complementary vector $Y(x_p^n)$. The algorithm is presented in Appendix B.

The relevance of each of these three methods will be evaluated in the following application Section, although when conditioning by a single value, only methods $m_1$ and $m_2$ are applicable.

Conditioning of Gaussian random fields

Conditioning by several points values. The usual method to generate Gaussian random fields conditioned by point values is described in several papers (Delhomme, 1979; Lantuejoul 1994). The process is illustrated in Fig. 3. It is based on the kriging of a field to which a simulated random field $Y*$ is added. This is processed in three steps:

1. Kriging the conditioning point values over the domain $D$ to obtain the field $Y$.
2. Simulating a kriging error field $\Sigma$.
3. Adding $\Sigma$ to $Y$ to obtain a Gaussian random field $Y_c$ conditioned by the point values.

The simulation of the kriging error field $\Sigma$ is achieved by:

1. Simulating a non-conditional random field $Y_{nc}(x)$.
2. Extracting from $Y_{nc}(x)$ the values at the conditioning point locations and kriging the extracted point values to obtain the field $Y_{nc}$.
3. Computing the difference between $Y_{nc}(x)$ and its kriging estimate $Y_{nc}^*$.

Finally the Gaussian field conditioned by the point values can be written as:

$$Y_c(x) = Y(x) + |Y_{nc}(x) - Y_{nc}^*(x)| \quad \text{for each } x \in D$$

Delhomme (1979) shows that the field $Y_c$ is a conditional simulation since (i) $Y_c$ is characterized by the same variability as the simulated variable $Y(x)$ (same cdf and spatial structure), (ii) it is merged with the conditioning points since the kriging is an exact
interpolator, meaning that the error at the conditioning station is equal to zero.

**Conditioning by a single point value.** Eq. (4) can be extended to the case of a single conditioning point by writing:

$$Y_c(x) = Y_{nc}(x) + \lambda(x)(Y(x_0) - Y_{nc}(x_0))$$

where \(x_0\) is the location of the conditioning point \(Y(x_0)\) and \(\lambda(x)\) is the weight coefficient of simple kriging written as:

$$\lambda(x) = \frac{C_y(x-x_0)}{C_y(0)}$$

if the spatial structure function is represented by the covariance function \(C_y\);

$$\lambda(x) = 1 - \gamma_y(x-x_0)$$

if the spatial structure function is represented by the variogram function \(\gamma_y\).

**Conditioning of non-Gaussian random fields by areal values**

A full description of the method used to condition the random fields by areal values can be found in Onibon et al. (2004). The method makes use of a Gibbs sampler (see Appendix B) coupled with an acceptation–rejection algorithm. As the other methods, the model requires the parameters of the cdf and the spatial structure function \(S_t\). Convergence considerations require the variance of the simulated field to be added as a supplementary model parameter.

**Application to the AMMA-CATCH Niger data set**

**Studied region, data and hydrological modeling**

The non-conditional and conditional algorithms are used to simulate the Sahelian rain fields observed by the AMMA-CATCH Niger (ACN) network (Fig. 4, see Cappelaere et al. (2009) for a comprehensive description of the ACN network) over the period between 1990 and 2002. The network consists of 30 tipping bucket rain gages providing rainfall data at a 5 min time-step. A total of 548 rainfall events were recorded over these 13 years (Balme et al., 2006). The simulation is carried out on a 60 x 60 km² sub-area where the density of rain gages is the highest. Fifteen rain gages have been selected inside and around this simulation area and used to (i) compute the statistical characteristics required as parameters of the rainfall stochastic model and to (ii) condition the simulations by point observations or estimated areal rainfall (see the modeling methodology shortly after).

![Fig. 3. Simulation process of Gaussian random fields conditioned by point values.](image-url)
Rain fields generated over this 60 × 60 km² area are used as forcing fields for the Dantiandou Kori hydrological system. The Hortonian model MSSM (Meso-Scale Surface Model) developed by Massuel (2005) is used to simulate the runoff on each of the 227 endorheic catchments of the Dantiandou Kori (Fig. 4). MSSM is based on a variation of the SCS equations (Soil Conservation Service, 1972) whose parameters are estimated for each endorheic catchment according to a set of physiographical characteristics. The model explicitly represents the Hortonian processes and requires rainfall inputs at 5 min time-step for event-wise simulations. Catchment sizes range from a few hectares to 50 km² and the global system covers a total area of 1200 km².

As may be seen in Fig. 4, the rainfall simulation domain is shifted in space with respect to the runoff modeling domain. This was done in order to ensure the most detailed possible documentation of the rainfall variability, with no significant impact expected regarding the relevance of the whole simulation process since: (i) rainfall characteristics have been shown to be stationary in space at the event scale over this region (Ali et al., 2003); (ii) the purpose of this hydrological modeling application is not to compare the output of each simulation procedure to observed runoff but rather to intercompare them in term of their statistical characteristics as explained later in the paper.

Modeling methodology

Rainfall modeling. Different types of rain fields are created.

- Kriged rain fields (referred hereafter to as kri): simple interpolation of the point rain gauges measurements (no simulation involved).
- Non-conditional rain fields: the simulations use either the mean event rainfall distribution parameters averaged over the whole period 1990–2002 (denoted nc simulations) or the mean event rainfall distribution characteristics averaged for each month of the rainy season (April–October) over the period 1990–2002 (nc-month).
- Rain fields conditioned by the observations at the 15 stations: the three methods \( m_1, m_2, m_3 \) presented in the previous section are tested (the simulations being, respectively, referred hereafter to as \( c_{15-m_1}, c_{15-m_2}, c_{15-m_3} \)).
- Rain fields conditioned by a single station (\( c_1 \)): due to its central position (Fig. 4) and the lowest rate of missing data (only three non-recorded events over the 13 years), the Kollo station is used as the conditioning station. Because of the three missing events at the Kollo station, only 545 of the 548 recorded events are simulated for all the simulation types.
- Finally the rain fields conditioned by areal values (\( c_A \)) are simulated.

The simulated series of event rain fields will be evaluated in the following section by analyzing the characteristics of simulated series of rain fields: mean event characteristics, mean seasonal cycle and inter-annual variability.

Runoff modeling. The different types of rain fields are used to force the hydrological model. As it requires 5 min time-step rainfall inputs, the event rain fields are temporally disaggregated according to the synthetic hyetograph proposed by Balme et al. (2006).

Implementation of the rain field model

Point rainfall distribution

The determination of the distribution of the simulated variable is required to implement the rainfall model. Previous studies (e.g. Balme et al., 2006) have shown that the event rain depth cdf can be represented by the mixture of a gamma distribution \( G \) and an atom at zero (frequency \( F_0 \)). The distribution function can be written as:

\[
F(x) = \begin{cases} 
F_0 & \text{if } x = 0 \\
G(x) & \text{otherwise}
\end{cases}
\]
The two parameters of the gamma function $G$ are:

$$
F(h) = \begin{cases} 
F_0 & \text{if } h = 0 \\
F_0 + (1 - F_0)G(h) & \text{if } h > 0
\end{cases}
$$ (8)

The two parameters of the gamma function $G$ are:

$$
p = \frac{E_0}{\text{Var}_0} \quad \text{(scale parameter)}
$$
and

$$
\lambda = \frac{\text{Var}_0}{E_0} \quad \text{(shape parameter)}
$$

where $E_0$ and $\text{Var}_0$ are the mean and variance of the event depth conditional upon it being non-zero. $F_0$, $E_0$ and $\text{Var}_0$ are thus the three parameters that characterize the distribution of the point rainfall in the area. The values of these three parameters were computed from the 545 events and are given in Table 1. Table 1 contains the mean values computed from the 15 selected stations (used to perform the simulations $nc$, $c_{15}$, $c_1$ and $c_6$) and the mean monthly values (used to perform the simulations $nc$-month). As a basis for commentaries, the values computed for the Kollo station are also reported, as well as the mean values computed by Balme et al. (2006) over the same period from the entire ACN network. The number of events, the non-conditional mean $E$ and variance $\text{Var}$ are also given.

The comparison between the values computed from the 30 stations and those computed from the 15 selected stations confirms the relative spatial stationarity of the parameters $E_0$ and $\text{Var}_0$ on the ACN network, already pointed out by Ali et al. (2003) and Balme et al. (2006). There are of course some local differences as shown by the low values of $E_0$ and $\text{Var}_0$ computed at the Kollo station. These values are the lowest among those computed for the 30 stations over the period 1990–2002. As expected, the value of $F_0$ is lower for the subset of 15 stations than for the total network. This highlights the fact that the number of events depends on the location of the stations, due to the criterion used for defining the events, which makes it more likely to observe events at the centre rather than on the border of the network window. The monthly values computed on the 15 stations illustrate the seasonality of the rainfall event characteristics. In particular, the high values of $E_0$ and $\text{Var}_0$, combined with low values of $F_0$ for the months of July and August result from the presence of big, intense and low intermittent OCSs during the core of the rainy season.

Spatial structure function

The spatial structure of rain fields is represented by the variogram proposed by Ali et al. (2005) from the analysis of the ACN data. It is a nested anisotropic structure defined as:

$$
\gamma(h) = 100 \left[ -\exp \left( -\frac{|h_1|}{20} \right) \right] + 105 \left[ 1 - \exp \left( -\frac{|h_2|}{200} \right) \right]
$$ (11a)

where $a_1$ is the anisotropy coefficient ($a_1 = 0.6, a_2 = 0.5$). The way the variogram used to simulate the Gaussian random fields was derived from the variogram above is explained in Appendix A.

Other modeling features

The fields are simulated on a 5 km resolution grid, which is fine enough according to the threshold resolution of 20 km suggested by Vischel and Lebel (2007) for hydrological modeling applications over the region.

In order to take into account the dispersion of the stochastic simulations, 100 sets of the series of 545 events were simulated for $nc$, $c_{15}$, $c_1$ and $c_6$. The kriging interpolation produces a single rain field for each event, which is smoother than the various simulations (the interpolated kriged rain field is in expectation the average of the $c_{15}$ rain fields).

To limit potential border effects, all of the 15 selected stations, even those outside the boundaries of the simulation area, were used to condition the kri and $c_{15}$ methods.

For the simulations $c_6$, the mean spatial event rain depths were estimated by kriging. As already mentioned, the field variance must be prescribed to the model. As visible from Fig. 5 the areal average and the standard deviation of the rain fields are correlated. For each simulated event, the standard deviation of the field is thus randomly assigned according to a Gaussian function defined from the graph of Fig. 5 for different classes of mean areal rainfall amount.

Evaluation of the stochastic model of rain fields

Retrieval of the event rain field characteristics

Spatial structure function

Fig. 6 shows the mean variogram averaged from the 100 sets of each type of simulation. The simulated variograms compare well with the empirical variogram. It can just be noted that the variogram of the $c_6$ simulations slightly overestimates the variance for the high inter-distances. Since the rainfall values are respected at the conditioning station locations, the observed variogram is naturally retrieved by the simulations $c_{15}$: the three methods $m_1$, $m_2$ and $m_3$ give similar variograms, thus only one variogram (labeled as $c_{15}$) has been reported in Fig. 6.

Frequency distributions

The three parameters of the cdf ($F_0$, $E_0$ and $\text{Var}_0$) are computed from the 100 realizations of the 545 simulated fields at each simu-
characteristics, with just a 1% underestimation for E0 is 209.2 mm² at Kollo compared to a 15 station average of 229.2 mm².

For the c15 simulations, the truncation of the Gaussian distribution produced by method m1, is directly reflected by an underestimation of the mean zero value frequency F0, as well as an underestimation of E0 and Var0 (Table 2). Method m1 is thus not recommended. Both methods m2 and m3 give the right mean characteristics, with just a 1% underestimation for F0. Method m2 being simpler to implement and much less time-consuming than method m3, it will be preferably used for our case study. Consequently, in the following, only the results referring to this method m2 will be presented. The spatial pattern of the characteristics is directly linked to the values imposed by the 15 conditioning stations and is thus quite variable (Fig. 7).

The simulations cA are characterized by an unexpected non-uniform spatial distribution of F0, E0 and Var0, which are underestimated at the centre of the simulation grid and overestimated at the borders. This behaviour has a limited influence on the mean parameters: F0 and E0 are slightly underestimated (21% and 14.1 mm for the simulations against 23% and 14.7 mm for the observed averages over the 15 selected stations), while the values of Var0 are slightly overestimated (235 mm² against 229 mm²).

Dispersion of the simulations on the mean annual rainfall

Fig. 8 shows the mean annual rainfall values of the 13 simulated years over the simulation area. Except for the kriging interpolation, for which only one realization per event is simulated, there is a dispersion of the mean annual rainfall values due to the 100 simulated realizations of the series of 545 events. The mean, the standard deviation and the associated coefficient of variation of the simulated mean annual rainfall values are reported in Table 3.

The mean annual rainfall reference is 476 mm, that is the value obtained by kriging of the observations. According to this reference, the errors of simulated rainfall are all lower than 1% except for the simulations c1 that are conditioned by the Kollo station, whose mean annual rainfall is 440 mm. As a consequence the mean annual rainfall for the simulations c1 is 452 mm, which is 5% lower than the kriging reference value: expectedly, the conditioning station has a strong impact on the average statistics.

The dispersion of the histograms of Fig. 8 is synthesized by the values of the variation coefficient reported in Table 3. Without any conditioning, the simulations nc are logically the most dispersed (standard deviation of 20.2 mm, i.e. a coefficient of variation of more than 4%). As expected the simulations cA are the most constrained (CV = 0.2%), since they have to respect the mean areal rainfall values. The conditioning by multiple points c15 also strongly reduces the dispersion of the simulation (CV = 0.7%). It is finally worth noting that the conditioning by a single station value reduces the dispersion of the simulations by a factor 2 (CV = 2% to be compared to 4% for the simulations nc).

Simulation of the inter-annual rainfall variability and the mean seasonal cycle

The capacity of the simulations to reproduce the inter-annual variability and the mean seasonal cycle of rainfall is studied here.
Fig. 7. Maps of the mean characteristics of the point frequency distribution computed for the different types of rain fields.

Table 2
Simulated point rainfall statistics.

<table>
<thead>
<tr>
<th></th>
<th>April–May</th>
<th>June</th>
<th>July</th>
<th>August</th>
<th>September–October</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_0$ (mm)</td>
<td>11.3</td>
<td>6.3</td>
<td>10.2</td>
<td>11.6</td>
<td>12.9</td>
</tr>
<tr>
<td>$Var$ (mm²)</td>
<td>214.3</td>
<td>137.5</td>
<td>190.8</td>
<td>216.6</td>
<td>241.9</td>
</tr>
<tr>
<td>$E_0$ (mm)</td>
<td>14.7</td>
<td>9.9</td>
<td>14.3</td>
<td>14.9</td>
<td>16.0</td>
</tr>
<tr>
<td>$Var_0$ (mm²)</td>
<td>229.2</td>
<td>179.8</td>
<td>208.7</td>
<td>229.0</td>
<td>250.5</td>
</tr>
<tr>
<td>$F_0$ (%)</td>
<td>23</td>
<td>36</td>
<td>29</td>
<td>22</td>
<td>20</td>
</tr>
</tbody>
</table>

Conditional simulations

<table>
<thead>
<tr>
<th></th>
<th>$c_{15}-m_1$</th>
<th>$c_{15}-m_2$</th>
<th>$c_{15}-m_3$</th>
<th>$c_1$</th>
<th>$c_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E_0$ (mm)</td>
<td>11.7</td>
<td>11.4</td>
<td>11.4</td>
<td>10.8</td>
<td>11.1</td>
</tr>
<tr>
<td>$Var$ (mm²)</td>
<td>211.5</td>
<td>216.0</td>
<td>216.0</td>
<td>195.7</td>
<td>217.9</td>
</tr>
<tr>
<td>$E_0$ (mm)</td>
<td>14.0</td>
<td>14.7</td>
<td>14.7</td>
<td>14.2</td>
<td>14.1</td>
</tr>
<tr>
<td>$Var_0$ (mm²)</td>
<td>220.4</td>
<td>229.6</td>
<td>229.6</td>
<td>209.2</td>
<td>235.0</td>
</tr>
<tr>
<td>$F_0$ (%)</td>
<td>16</td>
<td>22</td>
<td>22</td>
<td>24</td>
<td>21</td>
</tr>
</tbody>
</table>

Both of them are linked to two complementary factors: (i) the number of rainfall events and (ii) the intensity of the rainfall events. Here the number of rainfall events is fixed (equal to the number of observed events: 545). The simulated intensity per event is thus the only component likely to produce dispersion in the inter-annual rainfall variability and in the shape of the seasonal cycle from one set of simulations to another and/or between the various methods. Fig. 9 shows the values of the annual rainfall values for the 13 years of the simulation period averaged over 100 realizations. Fig. 10 shows the mean seasonal cycle of the mean event rainfall depth (Fig. 10a) and the corresponding seasonal cycle of daily rainfall (Fig. 10b) computed by assuming that days without rainfall events are zero values.

By referring again to the values obtained by kriging (kri), it can be seen that the inter-annual variability of rainfall and the shape of the seasonal cycle are respected only if the simulations are sufficiently constrained by observations. Logically the inter-annual variability is almost perfectly recreated for the simulations $c_{15}$ ($r^2 = 0.98$) and $c_A$ ($r^2 = 1$) (Fig. 9). In Fig. 10 the seasonal signals relative to the simulations kri, $c_A$ and $c_{15}$ are logically quite similar, since the spatial value of rainfall is estimated with an equivalent precision. All three methods show quite well the separation between the pre-onset conditions (installation of the monsoon) and the core of the rainy season which follows at the end of June and is characterized by more intense events.

On the opposite, the values obtained from the nc simulations do not always match the reference kriged values. Fig. 9 shows that for some years, the annual rainfall is almost reproduced correctly (1994, 1997, 2001 and 2002) but in other cases the difference is large as for instance in 1998 with a strong 35% underestimation. The nc simulations are equivalent to a rainfall signal in which all the events would have the same intensity (equal to the mean value $E = 11.3$ mm, Table 2). Since on our sample the number of events explains only 26% of the variance of the annual rainfall, the poor capacity of the nc simulations to reproduce correctly the inter-annual variability is not astonishing. The use of monthly characteristics (nc-month) slightly modulates the shape of the event rainfall seasonal signal (Fig. 10a). This could have influenced the inter-annual variability however the monthly modulation is negligible compared to the stochastic dispersion of the simulations and the nc-month inter-annual signal was similar to the nc one (not shown in Fig. 9).

Regarding our capacity to reproduce the inter-annual variability, the most noticeable result is the result obtained with simulations $c_1$. The annual rainfall values are much better simulated by this method than when using the nc simulations, even though only one conditioning station is used. For instance, the 30% error on the annual rainfall of year 1998 with the nc simulations is reduced to 13% in the $c_1$ simulation. The signal of annual values obtained from the $c_1$ simulations is correlated with a determination coefficient of 0.9 with the reference inter-annual signal, against 0.26 for the nc simulations. Note also that the seasonal cycle (Fig. 10a) obtained by the $c_1$ simulations has a shape quite similar to the reference signal even though the event depths in the middle of the rainy season are lower. The important implication of this result is the following: with only one series of long term observations in the past, one is able to simulate rainfall grids over

<table>
<thead>
<tr>
<th>Mean annual rainfall</th>
<th>Average (mm)</th>
<th>Standard deviation (mm)</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kri</td>
<td>476</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nc</td>
<td>472</td>
<td>20.2</td>
<td>4.3</td>
</tr>
<tr>
<td>$c_1$</td>
<td>452</td>
<td>9.0</td>
<td>2.0</td>
</tr>
<tr>
<td>$c_{15}$</td>
<td>479</td>
<td>3.6</td>
<td>0.75</td>
</tr>
<tr>
<td>$c_A$</td>
<td>473</td>
<td>1.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Fig. 8. Mean annual rainfall estimated from the different types of simulated rain fields.

Fig. 9. Annual rainfall estimated from the different types of simulated fields.

regions whose area is in the order of a few thousands km²; these grids will provide a realistic simulation of the spatial pattern at the event scale and of the temporal distribution at the intra-seasonal and inter-annual scales.

An auxiliary comment on the seasonal signals of Fig. 10b is given in Appendix C.

Fig. 10. (a) Seasonal signal of the mean event rain depth; (b) seasonal signal of daily rainfall (days without events are considered as zero values). The dispersion (grey) corresponds to the 80% confidence interval. The curves have been obtained by (i) applying a moving average of length 16 days on the daily series computed for each year from the average of the 100 simulations (except for kri), (ii) averaging the obtained signals over the 13 years.

Application to the simulation of runoff over the Dantiandou Kori

The impact of using the different types of rain fields is studied by computing the mean annual runoff produced by the MSSM hydrological model on the Dantiandou Kori.
Table 4
Average, standard deviation and coefficient of variation of the simulated mean annual runoff.

<table>
<thead>
<tr>
<th></th>
<th>Average (mm)</th>
<th>Standard deviation (mm)</th>
<th>Coefficient of variation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>kri</td>
<td>16.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>nc</td>
<td>21.8</td>
<td>3.5</td>
<td>16</td>
</tr>
<tr>
<td>c₁</td>
<td>17.7</td>
<td>1.7</td>
<td>9.6</td>
</tr>
<tr>
<td>c₁₅</td>
<td>21.3</td>
<td>0.8</td>
<td>3.6</td>
</tr>
<tr>
<td>c₆</td>
<td>17.6</td>
<td>1.1</td>
<td>6.2</td>
</tr>
</tbody>
</table>

While the mean annual rainfall does not vary much for the various types of simulation (Table 3), the simulated annual runoff significantly differs depending on the type of forcing rain fields used. The dispersion is also strongly method dependant, the highest being for the nc simulations (CV = 16%) and the smallest for the c₁₅ simulations (CV = 3.6%) (Table 4).

The simulations nc and c₁₅ produce similar mean annual runoff values (respectively, 21.8 and 21.3 mm). The runoff obtained from simulations c₁ is 17% lower than the runoff modeled from simulations c₁₅ (Fig. 11). This behaviour is consistent with the mean characteristics of the simulated rain fields influenced by the conditioning Kollo station which has a slightly lower mean event intensity kᵢ₀ (see earlier comment on Fig. 7). As already shown in several studies (e.g. Michaud and Sorooshian, 1994; Winchell et al., 1998; Vischel and Lebel, 2007), the intensity of rainfall has a major influence on the production of Hortonian runoff. As a consequence an underestimation of the rainfall intensity can be non-linearly amplified in the process of runoff production. The fact that the modeled catchments are mainly grouped in the centre of the simulation window accentuates the effect of rainfall underestimation of simulations c₁ on the simulated runoff.

This last comment remains valid for the c₆ simulations for which the simulated mean runoff is also 17% lower than the c₁₅ mean runoff value, except that this difference is linked to a numerical bias mentioned earlier in the paper. The results show that the numerical bias has a non negligible impact on the simulated runoff, even though the runoff values remain in the range of the values obtained from the c₁ simulations.

An important result is the difference of more than 25% between the runoff obtained from the stochastic simulations c₁₅ and from the kriging interpolation kri (respectively, 21.3 mm against 16.2 mm). This difference illustrates the impact of accounting for the spatial variability of the rainfall intensities within the rain field used to force the hydrological model. As shown in Fig. 12 the kriged field is a smoothed representation of the spatial rainfall variability, while the simulated field restores the rainfall variability over the whole rain field producing a more erratic spatial pattern which is more realistic (even if not real) than the kriged field. The main advantage of conditional simulations over kriging interpolation is that its ability to reproduce both the mean spatial structure and the point distribution of the rain fields, does not depend on the density of an observation network. Despite the fact that there are no runoff observations available at the Dantiandou Kori scale, this simulation exercise confirms the importance (already pointed out in Vischel and Lebel, 2007) of representing the small scale spatial rainfall variability for modeling Hortonian runoff. One can thus quite confidently claim that the simulated fields are more relevant than the interpolated fields for hydrological modeling applications. Obviously, simulations conditioned by point values provide a suitable framework for evaluating the spatial dispersion associated with a given rain gage network, which is highly valuable when studying the propagation of the uncertainties associated with rainfall inputs into spatially distributed hydrological models.

Conclusion and discussion

A unified framework was proposed to simulate rain fields at high spatial resolution (typically 1 km) in order to test the sensitivity of runoff to various possible rainfall regime modifications in the Sahel. Different conditioning modes were compared, corresponding to various situations in terms of data availability: long series at one station, mesoscale networks such as the AMMA-CATCH observing system, areal estimates from models or satellites.

Main results

The first result is in the domain of the methodology: combining meta-Gaussian functions with a Gibbs sampling approach in a geostatistical framework proved to be an efficient way of simulating realistic event rain fields for all conditioning modes tested here, even though some unexpected spatial bias was detected for the areal conditioning mode.

The second main result is with respect to the conditioning with one station only, which allows retrieving with good reliability the rainfall variability at inter-annual and intra-seasonal scales: this is important for simulating small scale runoff in the past decades.

The third result is that point conditioning or areal conditioning can alternatively be used, depending on the type of forcing information available.

The fourth result relates to hydrological simulations: the conditional simulation framework provides a more realistic spectrum of possible rain fields – and thus of possible runoff – than standard downscaling techniques based on linear interpolation. This a key point indeed for assessing the uncertainties associated with forcing hydrological models with the various climatic scenarios that are especially dispersed in the tropical belt and West Africa in particular.

Model applications

The parameters used here to characterize the rainfall regime at the event scale are: the occurrence rates, the extension and fragmentation of the event rain fields (Fᵢ₀, cdf parameters, Fᵢ₀ and Varᵢ₀. Simulations of a large range of possible rainfall regime modifications can thus be carried out by varying separately or simultaneously these parameters, and then testing the impact on runoff (such tests have started and their results will be reported in a forthcoming paper). The main advantage of the method is its robustness, once a comprehensive preliminary statistical analysis has been carried out.
On the other hand, the main limitation of purely statistical models resides in the fact that they are calibrated on observations with a limited physical background, thus making the extension of the results to non observed situations difficult – i.e. same regions in the future or other regions. However extending the results from the Niger mesoscale site to another Sahelian area is relatively straightforward as long as the regions of interest have similar characteristics to those of the Niger site (smooth topography, regional gradients governed by the gradients of monsoon penetration, rainfall produced essentially by a limited number of large convective systems). The nature of the rainfall regime being similar over the whole Sahelian band (see Ali et al., 2005 stating that the rain events characteristics are relatively stationary over the Sahel), it is reasonable to assume similar types of event rainfall cdf and spatial structure function for the whole region, with possible local adaptation of their parameters that can be obtained from long term rainfall series at a few points and the operation of denser networks for a limited period of time, such as the AMMA-CATCH observing system. Looking to the future requires linking physical and dynamical patterns simulated by atmospheric models to the stochastic model parameters (occurrence rate, $F_0$, $F_0$, $Var_0$), in order to adapt the downscaling relationships to the climate states of the future as foreseen by GCM scenarios (e.g. Wilby et al., 1998).

Interactions with other AMMA investigations

While most of the work presented here relies on the long term AMMA-CATCH monitoring, the enhanced observing system deployed over the Niger site during the AMMA-EOP years (2005–2007) will benefit both the hydrological and meteorological communities. For one, radar data acquired in 2006 and 2007 can be used for a finer testing of the availability of our downscaling techniques to retrieve small scale rainfall structures of importance for runoff generation. Secondly, the installation of surface flux (Ramier et al., 2009) and soil moisture (Pellarin et al., 2009) stations, starting in 2005 and still in operation, will make possible to investigate the coherence between the measured fluxes, soil moisture and the simulated runoff in terms of water budget closure at the local scale. Thirdly, the fluctuations of the regional aquifer level will constitute the “ground truth” to verify the accuracy of the runoff simulations for the past decades according to the forcing rain fields used, that is obtained either through point conditioning of the statistical downscaling approach presented here, or through dynamical atmospheric outputs provided by the AMMA modeling community.

Acknowledgements

This research was funded by IRD in the framework of the AMMA-CATCH “ORE” program, as well as by the French ECCO-PNRH program (project “Eau et végétation au Niger”). Based on a French initiative, AMMA was built by an international scientific group and is currently funded by a large number of agencies, especially from France, UK, US and Africa. It has been the beneficiary of a major financial contribution from the European Community’s Sixth Framework Research Programme. Detailed information on scientific coordination and funding is available on the AMMA International web site http://www.amma-international.org. The authors would like to thank Pr. Peter Lamb and the anonymous second reviewer for their valuable comments that help improve the manuscript.

Appendix A: Spatial structure function $S_{st}$ for Gaussian random fields

In Eq. (1), the anamorphosis function $\phi$ transforming Gaussian to non-Gaussian random fields may also transform the spatial structure function $S_{st}$. It is thus necessary to define a suitable spatial structure function $S_{st}$ expected for the non-Gaussian fields. If the cdf of the variables $Z(x)$ is continuous, $S_{st}$ can be analytically deduced from $S_{st}$. If the cdf of the variables $Z(x)$ is discontinuous (i.e. presenting accumulations of values or atoms), $S_{st}$ has to be adjusted by a trial and error method in order to retrieve $S_{st}$ after transformation by the anamorphosis function (Guillot, 1999).

In the case of Sahelian rain fields, for which the cdf presents an accumulation of zero values, Guillot (1999) showed that a good approximation of $S_{st}$ could be obtained by normalizing $S_{st}$ by its variance: $S_{st} \sim S_{st}/\sigma^2$. In section “Implementation of the rain field model”, the variogram of Eqs. (11a) and (11b) (defining the spatial structure $S_{st}$) was thus normalized and the obtained $S_{st}$ spatial structure was used to simulate the Gaussian fields required to simulate the fields $nc$, $c_{15}$ and $c_1$. Note that $S_{st}$ is also used in the Gibbs algorithm of the simulations $c_{t5}$-$m_5$. For the simulations $c_A$, the algorithm does not require a modification of the spatial structure function. $S_{st}$ was thus directly used as it is (Onibon et al., 2004).

Fig. 12. Example of a kriged (left) and a simulated (right) event rain field conditioned by point values. The pattern of the kriged rain field is smooth while the pattern of the simulated rain fields is more erratic and is probably more realistic.
Appendix B: Gibbs sampling algorithm used in the simulations $c_{15-m_3}$

The simulations $c_{15-m_3}$ are based on the Gibbs sampling method, which is a MCMC (Monte Carlo Markov Chain) algorithm. It is used here to solve the problem of atoms in cdf occurring for instance at zero when dealing with event rainfall cdfs. The Gibbs sampling algorithm is described hereafter in the particular case of a non-Gaussian cdf with an atom at zero, but the algorithm can be generalized to any atom value:

1. **Initialization**
   - Let $Y^0(x_i)$ be a vector composed of the zero values arising from realizations of a non-Gaussian stochastic function observed at $p$ conditioning points and $Y^1(x_i)$ a vector composed of the non-zero values arising from the same realizations.
   - The values $Y^0(x_i)$ are randomly assigned in the interval $[-S_0, S_0]$, with $S_0 = N^{-1/3}f_0$ where $N$ is the Gaussian distribution function and $f_0$ is the frequency of zero values.
   - The values of $Y^1(x_i)$ are assigned by applying the inverse of the anamorphosis function to the non-Gaussian conditioning points with non-zero values (Eq. (6)). The concatenation of $Y^0(x_i)$ and $Y^1(x_i)$ gives the vector $Y^0(x_i)$.

2. **Iterations**
   - **(a)** An initial value of $Y^0(x_i)$ is randomly selected.
   - **(b)** The distribution of the selected value is estimated conditionally to all the other values of the vector $Y^0(x_i)$; it is a Gaussian distribution with mean $\mu$ and variance $\sigma^2$ given by simple kriging of the other values (note that the spatial structure used here is the same as the one associated with the Gaussian values $S_0$ defined so that after anamorphosis the expected non-Gaussian spatial structure function $S_0$ is retrieved).
   - **(c)** A value is extracted from the conditional distribution defined in (b). This is achieved by randomly drawing a realization $u$ in a normal distribution until the associated variable $\mu + u \times \sigma^2$ is lower than $S_0$.
   - **(d)** The selected value in (a) is replaced by the computed value in (c) to obtain the vector $Y^1(x_i)$ and the vector $Y^0(x_i) = (Y^1(x_i), Y^0(x_i))$.
   - **(e)** Back to (a) and iterating until the value of iteration $i$ differs from the value of iteration $i - 1$ by less than an error criterion.

Appendix C: Auxiliary comment about seasonal cycles of Fig. 10b

The three simulations $k_{15}$ and $c_{4}$ give the typical shape of the seasonal rainfall as described in Sultan and Janicot (2003). The mean signal obtained from the simulations $nc$ is interesting since it gives the shape that would have the seasonal signal if all the events had a constant depth over the season. It isolates the sole effect of the event occurrences on the shape of the seasonal signal. In reference to the rainy season signal obtained from kriging, the signal obtained from the nc simulations explains 94% of the variance of the seasonal signal. This comparison shows that the shape of the rainfall seasonal signal is primarily determined by the timing of the rainfall events and is only secondarily modulated by the fluctuations of the mean event intensities within the year.

References


