



HAL
open science

Stochastorm: A Stochastic Rainfall Simulator for Convective Storms

Catherine Wilcox, Claire Aly, Théo Vischel, Gérémy Panthou, Juliette Blanchet, Guillaume Quantin, Thierry Lebel

► **To cite this version:**

Catherine Wilcox, Claire Aly, Théo Vischel, Gérémy Panthou, Juliette Blanchet, et al.. Stochastorm: A Stochastic Rainfall Simulator for Convective Storms. *Journal of Hydrometeorology*, 2021, 22 (2), pp.387-404. 10.1175/JHM-D-20-0017.1 . hal-03426615

HAL Id: hal-03426615

<https://hal.science/hal-03426615>

Submitted on 4 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Stochastorm: A Stochastic Rainfall Simulator for Convective Storms

CATHERINE WILCOX,^a CLAIRE ALY,^a THÉO VISCHEL,^a GÉRÉMY PANTHOU, JULIETTE BLANCHET,^a
GUILLAUME QUANTIN,^a AND THIERRY LEBEL^a

^a *Univ. Grenoble Alpes, IRD, CNRS, Grenoble INP, IGE, Grenoble, France*

(Manuscript received 22 January 2020, in final form 9 November 2020)

ABSTRACT: Stochastic rainfall generators aim to reproduce the main statistical features of rainfall at small spatial and temporal scales. The simulated synthetic rainfall series are recognized as suitable for use with impact analysis in water, agricultural, and ecological management. Convection-driven precipitation, dominant in certain regions of the world such as the intertropical belt regions, presents properties that require specific consideration when modeling: (i) strong rainfall intermittency, (ii) high variability of intensities within storms, (iii) strong spatiotemporal correlation of intensities, and (iv) marked seasonality of storm properties. In this article, improvements for an existing stochastic generator of rainfall fields that models convective storms are presented. Notable novelties include (i) the ability to model precipitation event timing, (ii) an improved temporal disaggregation scheme representing the rainfall distribution at subevent scales, and (iii) using covariates to reflect seasonal changes in precipitation occurrence and marginal distribution parameters. Extreme values are explicitly considered in the distribution of storm event intensities. The simulator is calibrated and validated using 28 years of 5-min precipitation data from the 30-rain-gauge AMMA-CATCH network in the Sahelian region of southwest Niger. Both large propagative systems and smaller local convective precipitation are generated. Results show that simulator improvements coherently represent the local climatology. The simulator can generate scenarios for impact studies with accurate representation of convective precipitation characteristics.

KEYWORDS: Convective storms/systems; Mesoscale systems; Precipitation; Stochastic models; Intraseasonal variability

1. Introduction

Stochastic rainfall generators (SRGs) aim to simulate realistic rainfall series by reproducing key statistical features of rainfall variability. As SRGs can generate long-term rainfall sequences at fine resolutions for a given climatology, they are useful in many applications, including risk assessment studies to estimate the return periods of rare events (Evin et al. 2018; Arnaud et al. 2016) and assessment of rainfall estimation uncertainties and their propagation into impact models (Renard et al. 2011; Borgomeo et al. 2014). These statistical models are complementary to physical atmospheric or climate models as they can be used in climate change impact studies as a means to downscale, correct bias, and disaggregate coarse-resolution climate model rainfall outputs (Wilks 2010; Peleg et al. 2015; Sørup et al. 2016; Peres and Cancelliere 2018). For these reasons, stochastic rainfall models are recognized as useful tools in numerous areas of environmental sciences for which rainfall is of major influence, for instance hydrology, agronomy, and ecology.

An aim of SRGs is to produce rainfall series at spatial and temporal resolutions compatible with those required as inputs to impact models. These resolutions can sometimes be less than 1 km in space and on the order of 1 min in time (e.g., Wilson et al. 1979; Troutman 1983). As SRGs are data driven, the development of increasingly accurate rain measurement instrumentation (from networks of recording rain gauges, radar data, satellite data) combined with dedicated statistical advances has enabled the

development of high-resolution stochastic simulations (e.g., Benoit et al. 2018a; Peleg et al. 2017; Guilloteau et al. 2018; Peleg et al. 2020).

At kilometeric and subdaily scales, SRGs attempt to capture the high intermittency of rainfall fields, the high variability of precipitation within rainy zones, and the dependence in time and space of rainfall fields. These characteristics form the common core of space–time SRGs. Existing statistical approaches can be divided into two main classes: multisite and random fields models. The first are an extension of single-site stochastic models over several distant locations (often corresponding to rain gauges). They are mainly based on nonparametric resampling methods; on parametric point processes based on the successive use of statistical rainfall occurrence and a statistical rainfall amount model; or on cluster point processes (Cowpertwait et al. 1996; Wilks 1998). The second category focuses on continuously simulating (on regular grids) the spatial variability of rainfall. This family includes rain cell models (Féral et al. 2003), scale invariance models (Serinaldi 2010; Lombardo et al. 2017; Raut et al. 2018), and meta-Gaussian random fields (Guillot 1999; Leblois and Creutin 2013; Benoit and Mariethoz 2017; Benoit et al. 2018a; Vaittinada Ayar et al. 2019).

Among the diversity of stochastic rainfall models, some challenges remain concerning the simulation of certain rainfall regime characteristics. The seasonality of rainfall is most often taken into account through parameter calibration by rain type (e.g., Benoit et al. 2018a), weather type (Vaittinada Ayar et al. 2019), month by month (e.g., Vischel et al. 2009; Paschalis et al. 2013; Peleg et al. 2017), or season by season, which often implies an arbitrary discretization of the seasonal continuity.

Corresponding author: Théo Vischel, theo.vischel@univ-grenoble-alpes.fr

DOI: 10.1175/JHM-D-20-0017.1

© 2021 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) (www.ametsoc.org/PUBSReuseLicenses).

Modeling spatial intermittency remains another issue. It is based either on an explicit separation of rainy and nonrainy simulated areas (e.g., [Leblois and Creutin 2013](#); [Paschalis et al. 2013](#)), which can sometime lead to abrupt discontinuities at dry/wet transitions (e.g., [Schleiss et al. 2014](#)), or it can be directly inserted as an accumulation of null values in the rain intensity marginal distribution which often leads to simulating only the most spatially consistent events ([Stehlik and Bárdossy 2002](#); [Guillot and Lebel 1999b](#); [Vischel et al. 2009](#)). Extreme rainfall has particular pertinence when conducting risk assessment. However, rainfall extremes are not explicitly taken into account in rainfall stochastic generators except in some rare studies (e.g., [Wilks and Wilby 1999](#); [Baxevani and Lennartsson 2015](#); [Evin et al. 2018](#)).

While major advances have been made allowing genuine flexibility in the high-resolution stochastic rainfall models ([Paschalis et al. 2013](#); [Peleg et al. 2017](#); [Benoit et al. 2018a,b](#)), their implementation often requires advanced measurement devices allowing accurate documentation of spatiotemporal dependencies of finescale rainfall fields and/or differentiation of rainfall types. A direct consequence is that the majority of the regions of application of high-resolution SRGs concern developed countries that have extensive measurement networks ([Breinl et al. 2017](#)).

Developing countries have two major obstacles that impede them from benefiting from high-resolution SRGs: (i) two-thirds of them are located in the intertropical belt where rainfall, mainly from convective storms, is very intermittent and variable in time and space on a very fine scale, and (ii) the lack of dense instrumentation networks for hydrometeorological monitoring limits the possibility of documenting rainfall characteristics at fine space–time scales.

In this manuscript, we propose a high spatial–temporal resolution (on the order of kilometers and minutes) SRG “Stochastorm” whose foundations have been motivated by the possibility to meet applications in developing countries. Although it is usable in other contexts, Stochastorm aims to

- (i) respond to the specificities of climates characterized by convective rainfall regimes, i.e., a succession of storms that produce highly intermittent and spatially correlated rainfall fields at fine spatiotemporal scales, and
- (ii) adapt to situations where high spatiotemporal resolution data are scarce, with limited possibilities of directly estimating finescale spatiotemporal structures or defining rainfall typologies in a systematic way.

Stochastorm derives its structure from the two constraints cited above by separating the simulation process into three stages of simulation: (i) the arrival of a storm, (ii) the simulation at high spatial resolution of cumulative rainfall fields under the trace of a storm, and (iii) the temporal disaggregation of rainfall fields at fine time steps based on a propagative hyetograph model.

The dissociation of the simulation at the event scale (storm accumulation scale, steps i and ii in the previous paragraph) and the intra-event simulation step iii aims at facilitating the estimation of the model parameters in a context of scarce rainfall data or data at time steps that only allow to estimate the

cumulative rainfall under the storm (typically daily or hourly data). It is based on the empirical hypothesis that the temporal disaggregation scheme, despite its simplicity, makes it possible to reproduce the main characteristics of rainfall at fine scales.

Stochastorm builds on an SRG initially proposed by [Lebel et al. \(1998\)](#). It includes several new features and statistical developments that significantly improve the latest version of [Vischel et al. \(2009\)](#). The originality of the approach lies first of all in the simulation of storm cumulative rainfall due to the integration of (i) the seasonality of rainfall simulated by the integration of temporal covariates that continuously reflect seasonal changes in the statistical distribution of occurrence and rainfall intensities; (ii) the intermittency of rainfall within the marginal distribution based on Gaussian fields with censored likelihood, which lift the limitation of simulating only the most consistent storms; and (iii) the extreme values in the marginal distribution of storm totals by placing themselves within the framework of extreme value theory for the marginal distribution. Stochastorm is in this sense an original contribution to the family of meta-Gaussian SRGs.

The paper also proposes a temporal disaggregation framework specific to propagative storms, including a convective and stratiform trailing structure whose internal separation is sometimes blurred in the absence of ground-based measurements and/or weather radar data. Although specific to convective storms, the approach and statistical developments proposed to describe the temporal disaggregation pattern can be adapted to other types of rainfall.

The objective of the paper is therefore to describe the methodological specificities implemented within Stochastorm to simulate rain fields at high spatiotemporal resolutions. Here, Stochastorm is evaluated in the Sahel. This region in West Africa is a relevant case study as its rainfall regime is characterized by highly variable and intermittent seasonal monsoon storms. We make use of a dataset from a network of 30 rain gauges covering an area of 10 000 km² in southeastern Niger and providing 5-min rainfall records over a 28-yr period (1990–2017) ([AMMA-CATCH 1990](#)). This dataset allows assessing Stochastorm’s capabilities and robustness to transcribe the different aspects of the rainfall regime in the Sahel, whether prescribed or not in the model, from event to subevent scales up to 5 min.

[Section 2](#) presents the model, particularly with regard to previous versions which provide the basis of the current model. [Section 3](#) presents the new developments for the current version. [Section 4](#) details how the model was implemented in the African Monsoon Multidisciplinary Analysis–Coupling the Tropical Atmosphere and the Hydrological Cycle (AMMA-CATCH) Niger study site. [Section 5](#) presents the results of model calibration and simulation. Finally, [section 6](#) presents our concluding remarks.

2. Stochastorm presentation

The main principles of the generator were initially proposed by [Lebel et al. \(1998\)](#) and further developed in [Guillot \(1999\)](#), [Guillot and Lebel \(1999a\)](#), [Balme et al. \(2006\)](#), [Onibon et al. \(2004\)](#), and [Vischel et al. \(2009\)](#). The above articles demonstrated

that the precipitation model globally represented both the spatial distribution and marginal distribution of precipitation events, which are important characteristics for evaluating hydrological impacts (Vischel et al. 2009).

The following section describes the rainfall simulator (Lebel et al. 1998; Guillot 1999; Vischel et al. 2009) in the state it was in before the most recent developments presented in section 3.

a. Occurrence

The previous rainfall simulator did not explicitly model when events occurred. An event simulation meant that a storm passed over the study/simulation window, but without a specific time or date assigned to the event. As rainfall event occurrence is a major characteristic of a given rainfall regime, there is a need to simulate it in the rainfall generator.

b. Event-based rain fields

An event-based rain field is defined here as the cumulative rainfall left by the passing of convective rainfall systems over the study site. The stochastic simulation of event-based rain fields is achieved within the framework of meta-Gaussian random functions (Benoit and Mariethoz 2017). Here, it consists of deriving non-Gaussian random fields from Gaussian random fields by using an anamorphosis function (Guillot 1999). Gaussian fields have well-known properties and can be generated more easily than spatial fields with other distributions (Emery 2002).

The transformation/anamorphosis of Gaussian to non-Gaussian fields is the main difficulty of the meta-Gaussian framework. In particular, the appropriate spatial structure of the random Gaussian fields must be prescribed in order to reproduce the expected spatial structure and marginal distribution of the non-Gaussian fields after anamorphosis. In case of discontinuous processes, like intermittent event-based rain fields which contain a mass of zero values in their marginal distribution, there is no analytical solution to assess the spatial structure function of Gaussian random fields (Guillot 1999; Guillot and Lebel 1999a; Emery 2002).

In the first version of the generator, Lebel et al. (1998) used the turning band method. Guillot and Lebel (1999b) implemented it with a nested anisotropic covariance function that displayed a more suitable simulation of the diversity of storms structures than in Lebel et al. (1998). Onibon et al. (2004) proposed to generate Gaussian fields by using the sequential method which, coupled with an acceptance-rejection algorithm, allows conditioning the simulations by surface average values. While an empirical trial and error approach was used in the first version, Vischel et al. (2009) proposed a Gibbs sampling algorithm to assess the spatial structure function of Gaussian fields represented by a nested anisotropic variogram.

Via the above methods, the marginal distribution and spatial structure of cumulative event rainfall amounts are obtained. The simulation outputs are punctual, with simulations recorded on locations determined by the user on a regular or irregular grid. The simulator as it was in Vischel et al. (2009) did not explicitly consider extreme values in the marginal distribution of cumulative event rainfall.

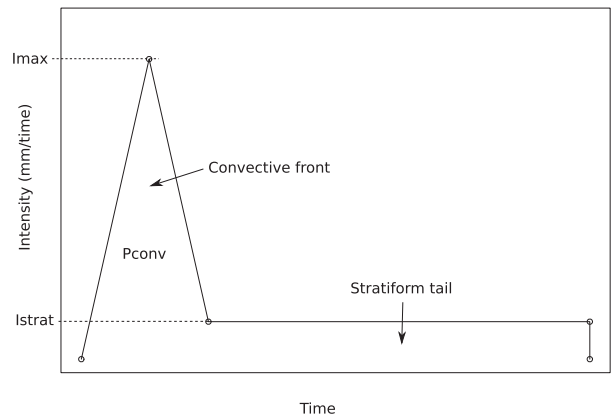


FIG. 1. Standard hyetograph shape from Balme et al. (2006) implemented in Stochastorm, consisting of convective and stratiform parts.

c. Temporal disaggregation

Temporal disaggregation, or simulating intensities at small time intervals within an event, is conducted with the aim of representing the physical properties of storms. This disaggregation is based on a deterministic synthetic hyetograph of a convective storm which consists of a symmetrical triangular peak representing the convective front of the storm, followed by a stratiform tail of lower intensity (Fig. 1). The hyetograph parameters (maximal intensity and duration) depend entirely on the total event rainfall via a relationship that evolved over the different model versions, the most recent published in Balme et al. (2006). The hyetograph shape can be adjusted provided that the morphological parameters of the hyetograph can be linked to the event accumulation to allow the temporal disaggregation of the cumulative event rainfall. The temporal disaggregation also includes a model of storm kinematics that consists of defining a field of hyetograph time of arrival based on prescribed storm propagation speed and direction.

3. New developments and technical definitions

The following sections detail the new features developed in Stochastorm. Table 1 provides an overview of model parameters.

Stochastorm has modules for adjusting model parameters to rainfall data which are described in the following sections. Note that in its current version, Stochastorm assumes that all parameters are homogeneous over the simulation area. The possibility to relax this assumption will be discussed in section 6.

a. Season limits and intraseasonal variability of parameters

The annual duration for which the rainfall regime is characterized by convective rainfall may be limited within the year, either because the rainfall typology changes along distinct seasons (as is the case in some temperate regions) or because the rainy season alternates with a dry season (as is the case in the intertropical belt regions). Modeling the start and end of

TABLE 1. Overview of model parameters. The asterisk symbol (*) indicates that the parameter is permitted to vary temporally over the season.

Category	Object	Parameters
Season definition	Start date	Normal distribution $\mathcal{N}(\mu_{\text{start}}, \sigma_{\text{start}})$
	End date	Normal distribution $\mathcal{N}(\mu_{\text{end}}, \sigma_{\text{end}})$
Occurrence	IET	Gamma distribution $F_\gamma(*b_{\text{IET}}, *k_{\text{IET}})$
Event rain field	Marginal distribution	$F_{\text{cumul}}(*p_0, *b_{\text{cumul}}, *k_{\text{cumul}}, u, \sigma_{\text{GPD}}, \xi_{\text{GPD}})$
	Spatial structure	Covariance function (ρ) and parameters (θ_ρ)
Temporal disaggregation	Propagation	Speed and direction
	Relationship I_{max}/P	Linear or beta distribution (min, max, α, β)

the wet season is thus an important feature to include in the occurrence model. Start and end dates of the rainy season are modeled in Stochastorm with a separate normal distribution for each.

Several parameters in Stochastorm are permitted to vary temporally over the rainy season according to the following equation:

$$\text{param}(t) = f_{\text{param}}(t, \boldsymbol{\theta}_{\text{param}}), \quad (1)$$

where t is the time covariate, f_{param} the selected function for representing seasonality, and $\boldsymbol{\theta}_{\text{param}}$ the vector of coefficients to be optimized for the function. The parameters that may vary in magnitude throughout the season are indicated in this article with the symbol *.

As an exploratory initial step, the parameters that may vary seasonally are first calibrated in a moving window over the data and then plotted to visualize their nonstationarity. From these plots, the types of possible functions for seasonal covariates are estimated visually. Then, statistical distributions using the different possible seasonal covariates are calibrated over the entire dataset. The covariate models are compared using AIC and likelihood ratio tests. The choice of seasonal covariates may also be informed by knowledge about local climatology in the study region.

Parameters are obtained by maximum likelihood estimation unless otherwise noted.

b. Event occurrence

The event occurrence in the new version Stochastorm is represented by the inter-event time (IET), which here is defined as the time between the start times of two events.

The probability density function (PDF) of the gamma distribution is denoted $f_\gamma(x, b, k)$ with a shape parameter $k > 0$ and a scale parameter $b > 0$:

$$f_\gamma(x, b, k) = \frac{1}{\Gamma(k)b^k} x^{k-1} e^{-x/b}. \quad (2)$$

The gamma distribution can alternatively be formulated using mean μ and shape k parameters given the relationship $\mu = kb$, or with shape k and rate r parameter given the relationship $r = 1/b$.

The IET parameter follows a gamma distribution with x as the IET value, modeled using the time-varying scale $*b_{\text{IET}}$ and shape $*k_{\text{IET}}$ parameters.

c. Marginal distribution of cumulative event rainfall

The marginal cumulative distribution (CDF) F_{cumul} of cumulative event rainfall at each station is defined by the following:

$$F_{\text{cumul}}(y) = \begin{cases} *p_0 + (1 - *p_0)F_\gamma(y, *b_{\text{cumul}}, *k_{\text{cumul}}), & \text{if } y < u, \\ p_u + (1 - p_u)F_{\text{GPD}}(y, \sigma_{\text{GPD}}, \xi_{\text{GPD}}), & \text{if } y \geq u, \end{cases} \quad (3)$$

where

$$p_u = *p_0 + (1 - *p_0)F_\gamma(u), \quad (4)$$

where y is the random variable of cumulative event rainfall at a given site; F_γ and F_{GPD} are the gamma and the GPD distributions, respectively; and $*p_0$ is the proportion of zero values in a spatial rain field. Events with cumulative rainfall below a given threshold u are modeled by a gamma distribution as in Eq. (2) with scale parameter $*b_{\text{cumul}}$ and shape parameter $*k_{\text{cumul}}$.

For events over the threshold u , the gamma distribution is replaced by a GPD distribution with a scale parameter σ_{GPD} and a shape parameter ξ_{GPD} . See [Baxevani and Lennartsson \(2015\)](#) and [Furrer and Katz \(2008\)](#) for similar approaches.

Since the gamma distribution only includes values below the threshold u , the parameters fitted by maximizing a censored likelihood (L^c) on positive data. Let \mathbf{y} be the matrix $(y_{ij})_{i=1, \dots, N; j=1, \dots, K}$ of the data, where i is the site (station) index, N the total number of sites, j the event index, and K the total number of events. The censored likelihood is then expressed as

$$L^c(\mathbf{y}, *b_{\text{cumul}}, *k_{\text{cumul}}) = \prod_{0 < y_{ij} < u} f_\gamma(y_{ij}, *b_{\text{cumul}}, *k_{\text{cumul}}) \times \prod_{y_{ij} \geq u} [1 - F_\gamma(u, *b_{\text{cumul}}, *k_{\text{cumul}})], \quad (5)$$

where F_γ is the CDF of the gamma distribution and f_γ is the PDF. Note that Eq. (5) considers that each site i is independent as well as each event j .

While seasonal covariates are technically possible for GPD parameters, it was chosen to keep the parameters stationary in time as uncertainty is already high and it would decrease the parsimony of the model. The sampling effect due to the rarity of extreme events limits the detection of a signal.

In this study, the GPD is fitted by maximum likelihood estimation on values larger than u .

d. *Spatial dependency: Gaussian fields with censored likelihood*

1) TRANSFORMATION OF DATA INTO CENSORED GAUSSIAN DATA

The spatial covariance structure determines how much the values at two different points vary together or not at a range of distances in a given direction. As a prerequisite to determining the spatial covariance structure, the non-Gaussian measured rain fields are first transformed into the process H . The process H is a version of the Gaussian process G but censored below a certain limit due to the zero values within the rain field.

Let Φ denote the $\mathcal{N}(0, 1)$ distribution. The following transformation is applied to the data for each event j ($j \in \{1, \dots, K\}$):

$$h_{ij} = \Phi^{-1}[F_{\text{cumul}_j}(y_{ij})], \tag{6}$$

where \mathbf{h} is the Gaussian version of \mathbf{y} . In this way, K (total number of events) realizations of a spatial process H are obtained. For each event j , the Gaussian field realization h_j is censored below the Gaussian value c_0 , corresponding to 0 mm in the observed rainfall field:

$$c_0 = \Phi^{-1}[F_{\text{cumul}_j}(0)], \tag{7}$$

where $F_{\text{cumul}_j}(0)$ is the equivalent of p_{0j} . The value c_0 is different for each realization due to the temporal dependence of F_{cumul} on t .

2) ESTIMATING THE COVARIANCE FUNCTION

To simulate the process H , the covariance function ρ of the process G must first be estimated. The Gibbs sampling algorithm proposed in the last model version (Vischel et al. 2009) presented numerical convergence problems, especially for marginal distributions where the frequency of zero values is high. Here, we propose an alternative approach based on censored likelihood that is better suited to elevated p_0 values.

Let g_{ij} denote the values of the realizations of the process G . Values of g_{ij} that exceed c_0 are known. For the values that do not exceed c_0 , it is known only that $g_{ij} \leq c_0$. For this reason, a classical likelihood calculation cannot be used to fit the model on data. Instead, a censored version of the likelihood is required.

Let \mathcal{D}_j be the set of sites where it does rain during the j th event:

$$\mathcal{D}_j = \{i \in \{1, N\} \text{ such that } g_{ij} > c_0\}. \tag{8}$$

Let θ_ρ be the parameters of the covariance function ρ that is to be estimated. Assuming that events are independent from one another, the censored likelihood of θ_ρ is

$$L^c(\theta_\rho) = \prod_j L_j^c(\theta_\rho). \tag{9}$$

If every g_{ij} is observed for the j th event (i.e., if it rains at all sites for the event j), then

$$L_j^c(\theta_\rho) = f_{\text{GP}}(g_{1j}, \dots, g_{Nj}; \theta_\rho), \tag{10}$$

where f_{GP} is the density of the Gaussian process with covariance ρ .

If at least one site is censored (within c_0), the likelihood becomes

$$L_j^c(\theta_\rho) = f_{\text{GP}}(\{g_{ij}\}_{i \in \mathcal{D}_j}; \theta_\rho) \Pr[\{G_{ij} \leq c_0(t_j)\}_{i \notin \mathcal{D}_j} | \{G_{ij} = g_{ij}\}_{i \in \mathcal{D}_j}; \theta_\rho]. \tag{11}$$

The probabilities are the CDF of a Gaussian process because

$$\{G_{ij}\}_{i \notin \mathcal{D}_j} | \{G_{ij} = g_{ij}\}_{i \in \mathcal{D}_j}, \tag{12}$$

is Gaussian with mean and variance given by

$$\mu = \Sigma_{1,2} \Sigma_{2,2}^{-1} \mathbf{g}_{\mathcal{D}_j}, \tag{13}$$

$$\text{var} = \Sigma_{1,1} - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{2,1}, \tag{14}$$

where $\mathbf{g}_{\mathcal{D}_j}$ is the vector formed by the elements of \mathcal{D}_j and Σ is the covariance matrix of the vector $(\mathbf{g}_{\mathcal{D}_j}, \mathbf{g}_{\mathcal{D}_j^c})$ with

$$\Sigma = \begin{pmatrix} \Sigma_{1,1} & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_{2,2} \end{pmatrix}. \tag{15}$$

The covariance ρ can be modeled with a variety of possible functions. Here an anisotropic covariance defined by a sum of exponential functions is fitted for rainfall events, given by

$$\rho(h) = v \exp\left(-\frac{\|h\|_{\text{anis}}}{\phi_1}\right) + (1 - v) \exp\left(-\frac{\|h\|_{\text{anis}}}{\phi_2}\right), \tag{16}$$

where v is the proportion of the field variance associated with the first exponential, and ϕ_1 and ϕ_2 are the range parameters of the two structures.

The notation $\|\cdot\|_{\text{anis}}$ denotes an anisotropic distance defined as

$$\|h\|_{\text{anis}}^2 = \|s_m - s_n\|_{\text{anis}}^2 = (s_m - s_n)' \mathbf{M} (s_m - s_n) \tag{17}$$

where

$$\mathbf{M} = \begin{pmatrix} \cos\psi & \sin\psi \\ -a \sin\psi & a \cos\psi \end{pmatrix},$$

with $a > 1$ and $\psi \in [-\pi/2; \pi/2]$.

Parameter ψ is the angle between the x axis and the major axis of the ellipse, and a is the ratio of the major axis length on the minor axis length. The level lines of the covariance function have an elliptical form.

See section 4c(4) for how the covariance structure was implemented in the present study.

e. *Temporal disaggregation*

The temporal disaggregation step takes the cumulative rainfall amounts and divides them into subevent intensities. The arrival time of the storm at a given point is conducted as in previous versions of the model (Guillot and Lebel 1999a),

Stochastorm simulation procedure

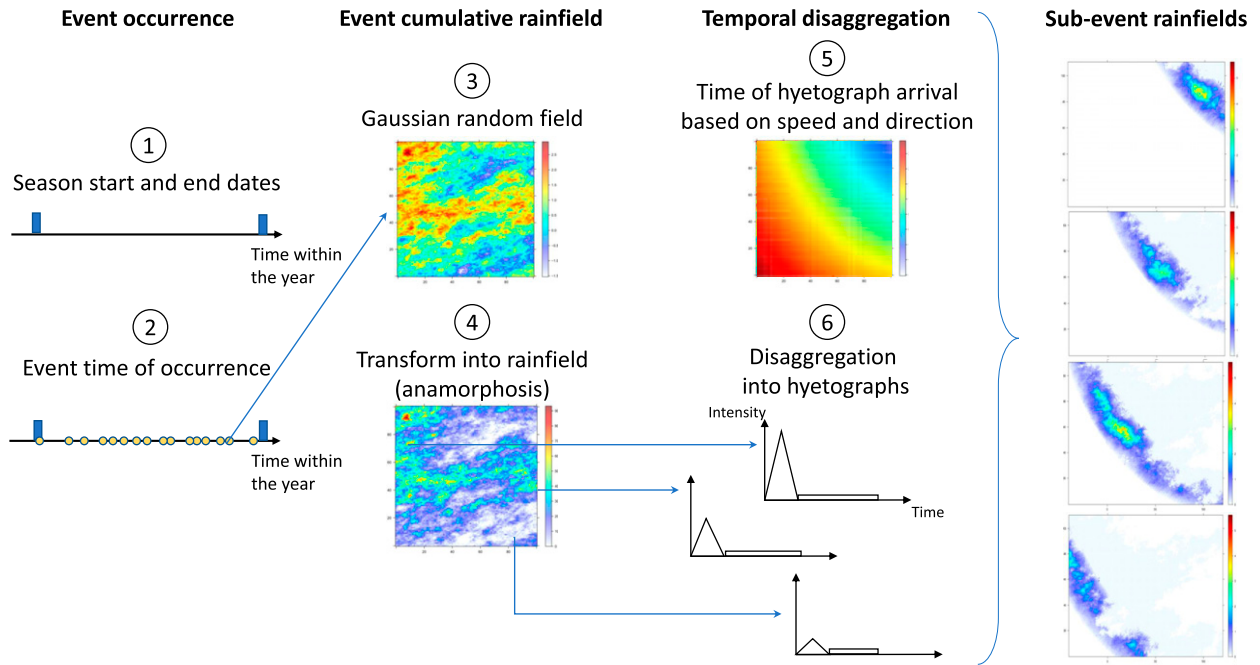


FIG. 2. Overview of simulation procedure for the Stochastorm model.

with a predetermined propagation speed and direction for each event based on empirical data.

Three variables are used to disaggregate the total rainfall at a site P into the synthetic hyetograph in Fig. 1: the convective portion of rainfall P_{conv} , the intensity of the stratiform tail I_{strat} , and the maximum intensity I_{max} . Balme et al. (2006) proposed the following relationships:

$$P_{conv} = \hat{c}_1 P + \hat{c}_2, \tag{18}$$

$$I_{strat} = \hat{c}_3, \tag{19}$$

$$I_{max} = \hat{c}_4 P + \hat{c}_5. \tag{20}$$

An alternative way of determining I_{max} instead of Eq. (20) is evaluated and compared in this study: A four-parameter beta distribution (two parameters from the standard beta distribution plus maximum and minimum bounds) with covariates relating I_{max} to P . This theoretically makes the relationship between the variables more flexible and more representative of the variability present within the observed rainfall systems.

The four-parameter beta PDF is defined as follows:

$$f_{beta}(I_{max}, \min, \max, \alpha, \beta) = \frac{(I_{max} - \min)^{\alpha-1} (\max - I_{max})^{\beta-1}}{\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)} (\max - \min)^{\alpha + \beta - 1}}, \tag{21}$$

where I_{max} is the maximum event rainfall intensity during an event, \min is the lower bound of the distribution, \max is the

upper bound of the distribution, and α and β are the two shape parameters. All four parameters can potentially have cumulative event rainfall as a covariate.

f. Simulation procedure

Once the parameters are obtained, the following simulation steps are conducted (see Fig. 2 for an overview):

- (i) Simulation of event start times.
 - 1) Wet season start and season end dates are generated from their respective calibrated normal distributions.
 - 2) Starting from the season start date simulated in the above step, dates of event starts are sequentially generated using the occurrence model of IET values (section 3b) until the end date of the season is exceeded. The last event is kept or discarded according to a Bernoulli distribution with $p = 0.5$.
- (ii) Simulation of cumulative event rainfall.
 - 3) One Gaussian field [marginal distribution $\sim \mathcal{N}(0, 1)$] is simulated per simulated event start.
 - 4) The marginal distribution $\mathcal{N}(0, 1)$ of the Gaussian fields is transformed (anamorphosis) to the established event marginal distribution F_{cumul} using the parameter values at the covariate t of the event date.
- (iii) Temporal disaggregation.
 - 5) A speed and direction is determined for the event and used to generate hyetograph time of arrival at each point.
 - 6) Cumulative event rainfall per site is disaggregated into the determined time step using the selected method for determining the I_{max} value of the hyetograph.

4. Application to Sahelian storms

a. Sahelian hydroclimatology

The Sahel is a band located roughly between the 250- and 750-mm isohyetal lines of the sub-Saharan precipitation gradient. Sahelian climate is driven by the West African monsoon, which determines the magnitude and frequency of monsoon storms and dictates a pronounced seasonal signal for the region (Lebel et al. 2003; D'Amato and Lebel 1998). The rainy season extends approximately from April to October, with peak rainfall months in July and August and little to no rainfall outside the rainy season. Monsoon storms are mainly derived from large organized mesoscale convective systems that generally propagate from east to west at speeds on the order of 30–50 km h⁻¹. Local storms with little or no propagation may also develop and represent less than 10% of the mean annual rainfall (Mathon et al. 2002). The average cumulative amounts and variance of precipitation events are higher in the middle of the season, with a lower percentage of null values (i.e., more large storms in July–August) (Ali et al. 2003; Balme et al. 2006).

The consideration of extreme storms is of particular importance in the Sahel. An increase in extreme precipitation values has been documented (Taylor et al. 2017; Panthou et al. 2014, 2018). Given the recent series of catastrophic floods in the region (Descroix et al. 2012; Sighomnou et al. 2013; Wilcox et al. 2018), it is important to understand how changes in precipitation extremes may drive extreme hydrological events.

b. Data used: AMMA-CATCH network

The calibration and validation of the simulator is performed with a subset of the AMMA-CATCH network of recording rain gauges. AMMA-CATCH consists of ecohydrological data collection at three mesoscale sites in West Africa. It is unique for the region in both its spatial density and the length of continuous data recording, since 1990 in Niger, 1999 in Benin, and 2002 in Mali. The data are freely available from the AMMA-CATCH database (AMMA-CATCH 1990).

Stochastorm was calibrated on the AMMA-CATCH study site located in the Sahel near Niamey, Niger. The site features a set of 30 rain gauges located within a 120 × 160 km² area (Fig. 3 providing 28 years of data (1990–2017) at the 5-min time step.

c. Model implementation

The following sections describe the implementation of the Stochastorm model on the AMMA-CATCH dataset located within the West African Sahel. Previous studies have shown that the assumption of spatial stationarity is valid for our study site [Ali et al. (2003) for the intermittence of rainfall events; Vischel et al. (2009) for the marginal distribution (gamma); Panthou et al. (2014) for the GPD distribution of extreme values].

Table 2 provides a summary of parameter values.

1) EVENT DEFINITIONS

An event consists of the rainfall over the study window as an individual storm passes. A rainfall event is considered to be distinct when it is separated in time from other series of precipitation; here, we define an event as being separated by at

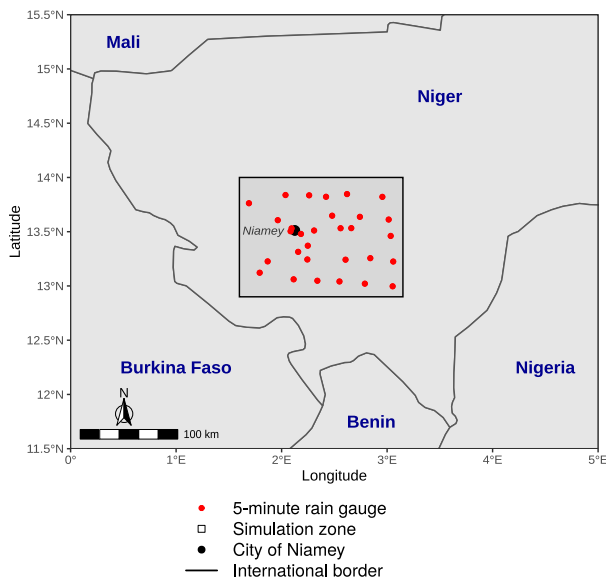


FIG. 3. Location of automatic rain gauge stations for the AMMA-CATCH network near Niamey, Niger.

least 30 min of no precipitation at the station level, with at least one time step (5 min) of no rain in the entire window of study.

The application of the previous version of the simulator in Vischel et al. (2009) only considered the most consistent rainfall events associated with organized and propagative mesoscale convective systems. As a consequence, the simulations were missing nearly 10% of the annual cumulative rainfall associated with smaller and less propagative local convection events. These small events are characterized by a strong intermittency (and thus a high frequency of zero values in the event rainfall marginal distribution) that, in practice, limits the possibility of convergence of the Gibbs sampling method to transform rainfall into Gaussian values in the anamorphosis process.

Here we benefit from the use of the censored likelihood approach that is much more efficient in handling zero values. While Vischel et al. (2009) only selected rainfall events covering at least 10/30 of the AMMA-CATCH Niger rain gauge network, here we define events as having at least 2/30 (or an equivalent percentage) of the rain gauge network receive more than 1 mm of cumulative rainfall.

Smaller events that did not meet the criterion were considered inconsequential; they may be due to erroneous data (e.g., a false tipping bucket), or precipitation that is too sparse to be organized into a storm. Measurements from events that were not considered in the study contributed less than 0.01% of the annual rainfall amount.

Once the initial event subset was created, the events were then classed into “large” (propagative storms) and “small” (local convection) categories. A large event has 30% or more of stations/data points registering rainfall. All other events were labeled as small.

For the dataset used in this study, 64.6% of events are large and 35.4% are small. Large events produce on average 95.7% of the total cumulative annual rainfall.

TABLE 2. Equations for Stochastorm parameters calibrated on AMMA-CATCH 1990–2017 data. Parameter t is the start day of the event (used for parameters that are modeled as nonstationary throughout the season, indicated by * within the text of the article), and P represents the cumulative event rainfall amount for a given event.

Variable	Equation	Section
Season start date	$\mathcal{N}(\mu_{\text{start}}, \sigma_{\text{start}}) = \mathcal{N}(19.88, 11.69)$	Section 3a
Season end date	$\mathcal{N}(\mu_{\text{end}}, \sigma_{\text{end}}) = \mathcal{N}(199.61, 6.88)$	
IET	$k_{\text{IET}}(t) = 1.08 \exp[-2.51(10^{-4})(t - 124.93)^2] + 0.66$ $\mu_{\text{IET}}(t) = 8.55 - 0.11t + 0.00046t^2$	Section 3b
Marginal distribution, large events	$p_0(t) = 0.47 - 3.20(10^{-4})t - 3.33(10^{-5})t^2 + 1.68(10^{-7})t^3$ $k_{\text{cumul}} = 0.86$ $r_{\text{cumul}}(t) = 0.15 - 1.65(10^{-3})t + 6.79(10^{-6})t^2$ $u = 40 \text{ mm}$ $\sigma_{\text{GPD}} = 13.40$ $\xi_{\text{GPD}} = 0.12$	Section 3c
Marginal distribution, small events	$p_0 = 0.85$ $k_{\text{cumul}} = 0.82$ $r_{\text{cumul}} = 0.16$	Section 4c(3)
Spatial structure	$\theta_{\rho, \text{large}}: \phi_1 = 9.91, \phi_2 = 152.74, \nu = 0.13, \psi = 0.28, b = 2.07$ $\theta_{\rho, \text{small}}: \phi = 30.89, \nu = 1$	Section 4c(4)
Base temporal disaggregation	$\hat{c}_1 = 0.89, \hat{c}_2 = -0.08, \hat{c}_3 = 2$ $\hat{c}_4 = 2.01, \hat{c}_5 = 0.53 \text{ (mm h}^{-1}\text{)}$	Section 3e
Relationship I_{max}/P	$\alpha = 0.30 + 1.07 \log(P)$	Section 3e
Beta distribution	$\beta = 36.41 - 36.46 \exp(-0.021P)$	

2) SEASON DEFINITION AND OCCURRENCE

Season limits are defined from April to October, a range that captures the monsoon rainfall which primarily occurs between June and September.

The seasonal covariate t is defined as follows: For central Sahel, the season extends from 1 April ($t = 1$) to 31 October ($t = 215$), with t being a decimal number. The integer digit of this decimal indicates the day in the season of the event start and the decimal part corresponds to the hour and minutes of the event start.

For the Sahel, we would expect parameters that are nonstationary over the rainy season, producing smaller IET values (i.e., more frequent storms) in the middle of the season (July–August).

The distribution of large and small events [section 4c(3)] throughout the season is modeled empirically with a Bernoulli distribution based on the input data. A spline is used to represent the proportion of small events throughout the season (Fig. 4). The value of the spline at a given date provides the parameter of the Bernoulli distribution at that point in the season. For simulation, this Bernoulli distribution is used to randomly generate large and small event labels at each simulated event start day.

3) MARGINAL DISTRIBUTIONS OF CUMULATIVE EVENT RAINFALL

Each category (large/small) has its own set of associated parameters for the marginal distribution of cumulative event rainfall. The large storms are modeled with a gamma + GPD distribution as described in section 3c, whereas small events are modeled only by a gamma distribution. Large event gamma distribution parameters were permitted

to vary over the season, as the nonstationary model had better fit than the stationary model. The gamma distribution for small events was modeled as constant over the season due to the relatively small variability compared to large events.

In this study, the threshold u for the GPD distribution is fixed at 40 mm of rainfall per event, approximately the quantile 0.97 as defined in Blanchet et al. (2018).

For the Sahel, we anticipate that the marginal distribution parameters for cumulative rainfall values and p_0 to have a seasonal evolution that leads to larger events (in both size and magnitude) in the middle of the season.

4) COVARIANCE FUNCTIONS USED TO MODEL SPATIAL DEPENDENCY

The methods described in section 3d are used to fit covariance functions on series of small and large events. There is one

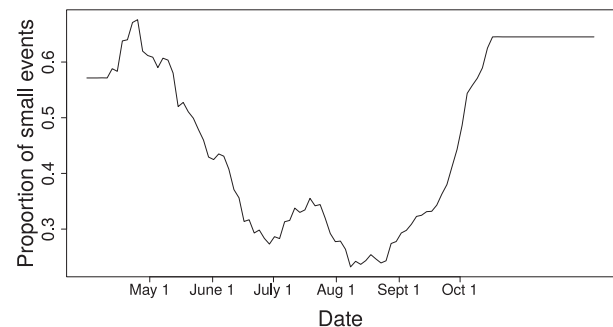


FIG. 4. The proportion of small events throughout the season for the AMMA-CATCH 1990–2017 dataset.

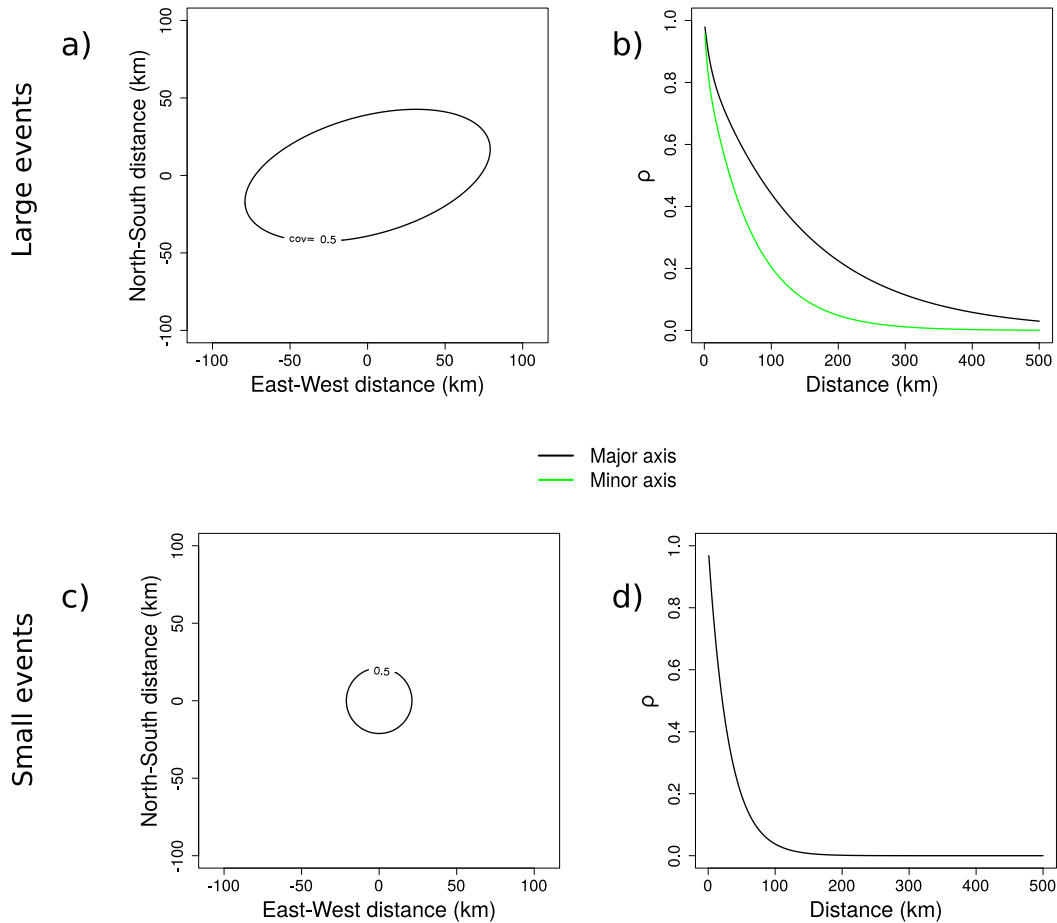


FIG. 5. Modeled covariance structure for (a),(b) large events and (c),(d) small events. Panels (a) and (c) display the contour lines of the distances at which $\rho = 0.5$. Panel (b) shows the evolution of ρ over distance for the major and minor axes for large events, and (d) shows the evolution of ρ over distance for any modeled direction.

spatial dependency structure for small events and one spatial dependency structure for large events.

The covariance structure for large events is given in Eq. (16). For small events, ν in Eq. (16) is set to 1 which simplifies the structure to an isotropic exponential covariance function.

Figure 5 displays the results for the modeled spatial structure of precipitation. See Table 2 for the specific parameter values after calibration on AMMA-CATCH data.

5) TEMPORAL DISAGGREGATION

The synthetic hyetograph [Fig. 1, Eqs. (18)–(20)] from Balme et al. (2006) was implemented in the simulator (see values in Table 2).

The present study aims to improve the estimation of the maximum intensity I_{max} of the synthetic hyetograph via the use of a four-parameter beta distribution [Eq. (21)]. An initial exploratory analysis was conducted by calibrating beta distribution parameters on I_{max} over a moving window by rainfall amounts from AMMA-CATCH data (e.g., I_{max} values sorted by magnitude of corresponding rainfall amount). The analysis revealed strong evidence of nonstationarity, with all four parameters increasing as rainfall amount increased.

The four parameters of the beta distribution in Eq. (21) were thus determined to depend on cumulative event precipitation.

A function of best fit was determined for each parameter. The beta distribution upper bound (max) was modeled using a two-part function (Fig. 6). For a lower range of cumulative rainfall values, the total event rainfall was used as the value of max. For larger values, max was defined as a log function that must exceed 2 times the observed I_{max} values. The switch between the two types of bounds was identified as the intersection between the cumulative rainfall value and the log function. We consider this to be a reasonable limit as the highest max values are near the world record for 5-min precipitation intensities (Burt 2007).

For the lower bound (min), linear segments that pass beneath the lowest observed values were used for smaller rainfall values, then extrapolated with a log function at the same point the log function starts for the max (Fig. 6).

For the α and β parameters, a log function and an exponential function were chosen respectively to link their values with the covariate of cumulative event rainfall. See Table 2 for their specific equations.

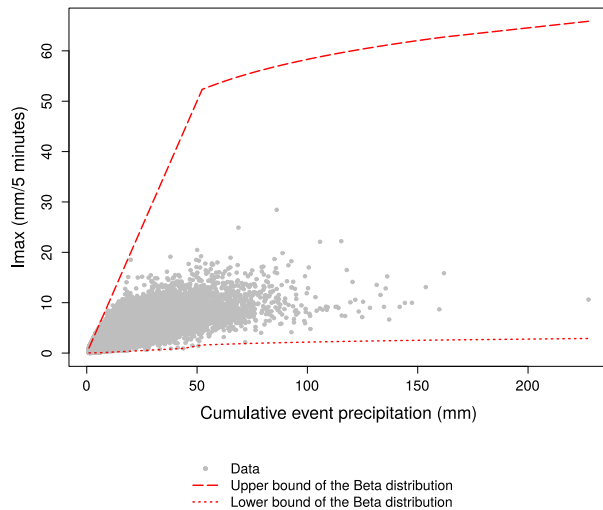


FIG. 6. Upper and lower bounds of the underlying beta distribution used to model maximum intensity I_{\max} with cumulative event rainfall as covariates for the beta distribution parameters.

Speed and direction are determined using an input database generated from AMMA-CATCH data. The recorded event with the value of cumulative event rainfall (P) closest to a given simulated cumulative rainfall amount is identified. Then, the speed and direction that was measured for the recorded event is used for the simulated event. This approach resembles the selection of an analog event within the recorded data used in, e.g., Chardon et al. (2014).

5. Evaluation of the model using AMMA-CATCH data (Sahel)

From Stochastorm calibrated on AMMA-CATCH data, we generated 30 realizations over a period of 28 years, equivalent to that of the dataset. The model permits the user to select the simulation sites; we chose to simulate at the locations of the 30 AMMA-CATCH measurement stations in order to compare simulation results with observed data.

The following sections describe the ability of the model simulations to reproduce both model parameters (i.e., whether simulated outputs reflect calibrated parameters) and observed characteristics of storms not directly prescribed in the model (whether simulated outputs are coherent with observations).

a. Event occurrence

Figure 7a displays the simulation results for the first and last days of the season. One can observe that both the model and simulations closely match the data, and that the first and last days of the season are well represented by a normal distribution. Note that although the season length was not explicitly parameterized in the Stochastorm model, simulations accurately represent the original data's season length (Fig. 7b).

Results for the IET (Fig. 8) also show good coherence between the simulation outputs and the original calibration data. Storms are reproduced at a frequency which matches that of the recorded record, with more frequent storms (lower IET

values) in July–September and less frequent storms at the beginning and end of the season. The main difference is toward the end of the season, where the seasonal evolution in the simulations appears to be underestimated relative to model parameters. This could be explained by the use of the Bernoulli law to decide whether or not to keep the final event of the season, and by the low frequency of events at this part of the season. The simulations remain relatively coherent with the values from rain gauge data.

Although the number of events per season was also not explicitly modeled, the number of simulated events generated from the IET parameter are representative of the data (Fig. 8). One can note however that the simulations have difficulty in retranscribing certain values, such as that for 1994. Considered as exceptional, notably because of an unusual occurrence of both small and large events, this year shows that extreme values of occurrence are not easy to reach by the proposed occurrence model. While the simulator is technically capable of producing a simulation with the amount of events seen in 1994, with a gamma distribution for IET values, it is unlikely to see such extreme cases very often in the simulations. The model is still able to replicate typical numbers of events and their variability.

b. Marginal distribution of rainfall

Figures 9a and 9b show the mean and variance of the cumulative event rainfall amounts and Fig. 9c shows the evolution of p_0 . The seasonal signals of mean values and variance are well represented.

Of note are the results for p_0 (Fig. 9c), the event spatial extent represented by the proportion of nonzero values. The simulations were conducted separately for large and small events. When the simulations are combined, the seasonal signal of p_0 in the observed data is accurately reproduced. The variability of the spatial extent of the events (as determined by the proportion of stations measuring rainfall during the event) is also well simulated (not shown). This is a notable result coming from the ability of the censored likelihood to simulate both large and small events. The arbitrary simulation of only large events (imposed in the previous version of the model) hindered the capacity of the simulator to reproduce the distribution of p_0 found in the data. With the revised model, the overall distribution of p_0 is reproduced.

Figure 10 shows the quantile–quantile plot of the simulated and observed rainfall values at event, hourly, and 5-min temporal scales. The simulation outputs have good coherence with the original data at both the event and hourly scales (Figs. 10a,b). For many events most rainfall occurs within the first hour (during the convective portion of the storm), which explains why the results for event and hourly scales are similar. In particular the extremes produced by the simulator fall closely along the reference line. The very highest quantile is somewhat underestimated (with the exception of an outlier) but realistic, and the quantiles below match the data well.

In Fig. 10c, one can see that the simulations slightly overestimate smaller values (approximately 1–4 mm per 5 min) and underestimate larger values. This is likely due to the hyetograph used for the disaggregation which smooths the rain intensities via its very simplified form. The bias, however,

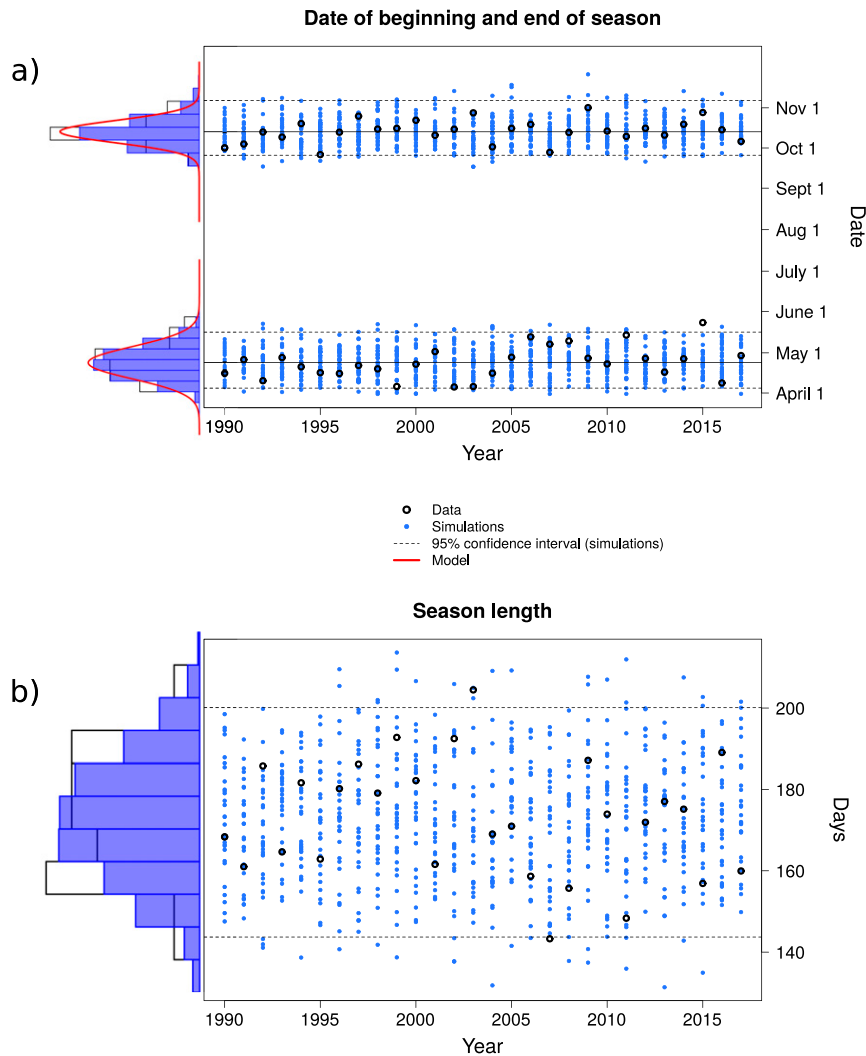


FIG. 7. (a) Distributions of beginning and end of season dates, with data, simulations, and model (see histogram to left of plot). (b) Season length (number of days from first day with precipitation to last day within the same year). See Table 2 for the specific parameter values of the model used to generate season start and end dates.

is relatively small, indicating that the methodology produces results that are close to realistic despite the simplicity of the disaggregation scheme.

c. Spatial structure

Figure 11 compares the observed and simulated empirical variograms at different time scales. For large events, the simulations show a slight overestimation, especially for longer distances. This could in part be explained due to the restricted spatial area of the study, which means there is a lack of observations at longer distances. The simulations are more coherent with the observations over the major axis than over the minor axis, which could be due to the more consistent organization of storms in the direction of propagation.

For small events, the values for the simulated variograms were lower than those of the observed variogram. Some of the

biases displayed could in part be due to the capacity of the model to simulate a small event using large event parameters and vice versa.

The results were consistent between the hourly and event time scales. At the 5-min time scale, the simulations underestimated the empirical variograms. This is again a mark of the simplistic shape of the hyetograph which does not represent all the variability of rainfall intensities at 5 min. This results in an underestimation of the variance of the variogram. The underestimation of the variance is however acceptable and the general pattern of spatial structure is maintained.

d. Robustness

To determine the robustness of the model, the data used for calibration was divided into two subsets: data from odd years and data from even years. In comparing the data, it was

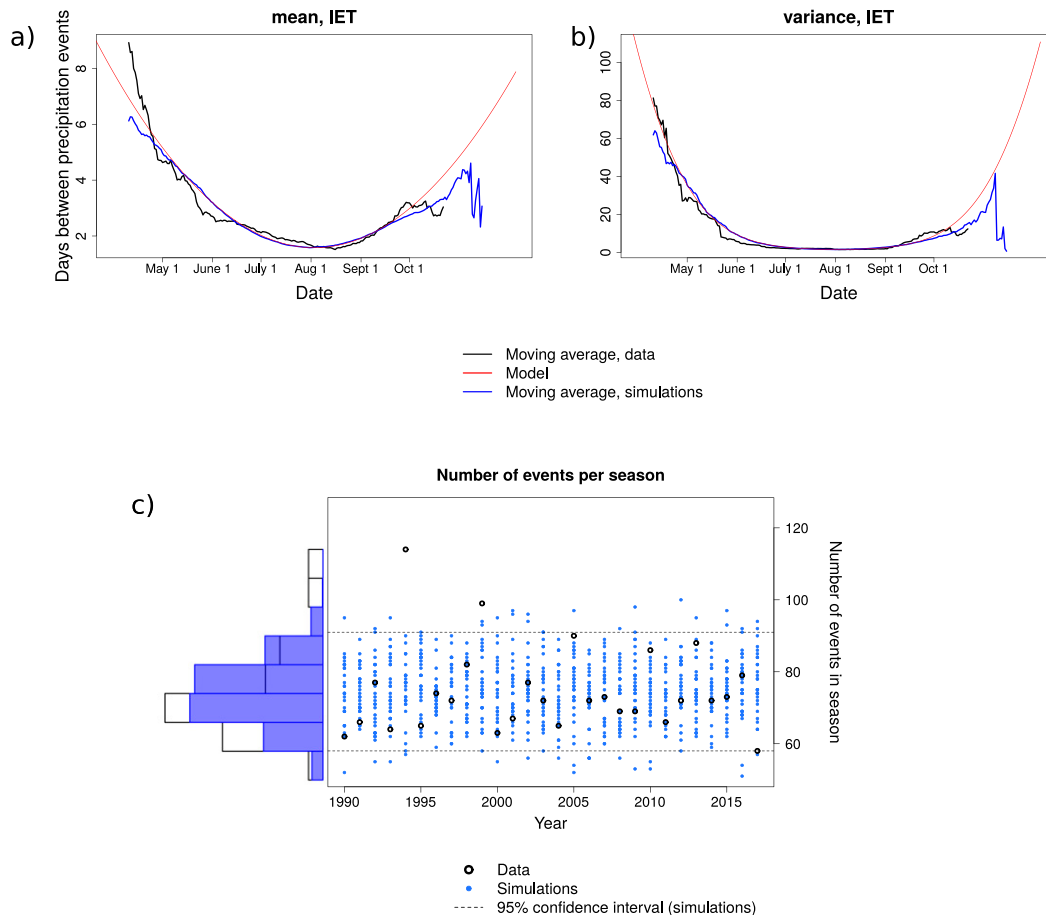


FIG. 8. (a),(b) Results for the IET and (c) number of precipitation events per season. See Table 2 for the specific parameter values and formulations of the model used for IET.

observed that essential features such as the seasonal rainfall signal were preserved for both the magnitude and frequency of events. Each dataset was then used to calibrate a set of parameters in Stochastorm and subsequently generate rainfall event simulations. Figure 12 shows a summary of this analysis. We did not find a significant difference in the simulations produced by each subset of data. This leads us to conclude that the model is robust enough to be calibrated on rainfall series of relatively short durations while keeping its simulations skills.

e. Hyetograph model evaluation

Figure 13 compares observed data with the maximum intensities generated from two methods of temporal disaggregation. The overestimation of maximum intensities for larger cumulative precipitation values when using the previous method is corrected when using the newly implemented four-parameter beta distribution (Fig. 13a). The beta distribution method more realistically models both the average and natural variability in maximum intensities; rather than having a set maximum intensity that corresponds to a given cumulative precipitation value as in the previous model, the maximum intensities are generated from the underlying distribution,

leading to a variety of magnitudes which correspond more to the behavior of the observations (Fig. 13b).

Figure 14 shows the hyetographs of events that have a given cumulative precipitation value. Although only the peak event intensity was specifically generated, the magnitudes of other subevent intensities were well reproduced by the temporal disaggregation method using the synthetic hyetograph. Both the average hyetograph and the variability (confidence interval) associated with a given cumulative value were reproduced. This also ensures a good coherence of the temporal autocorrelation of the hyetographs that we verified along different lags (not shown).

6. Discussion and conclusions

The above study presents the recent developments for the SRG Stochastorm, a modeling and simulation tool that aims to replicate the properties of convective storms. The simulator can generate nonstationary intra-annual precipitation scenarios at fine spatiotemporal scales.

Features include modeling of event occurrence dates via the inter-event time (IET) parameter and simulation of rainy season start and end dates. In addition, the model includes:

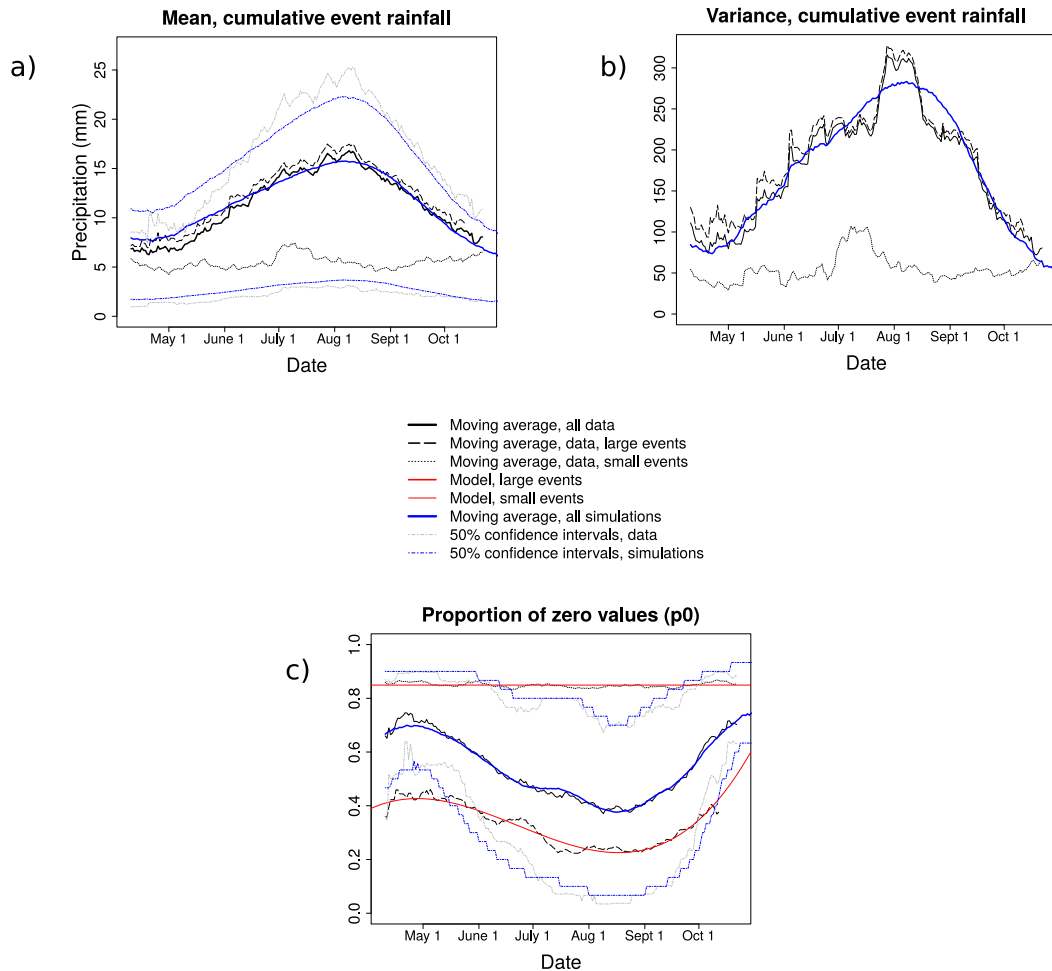


FIG. 9. Results for cumulative event rainfall over the study area (30 stations/simulation points aggregated) with 50% confidence intervals. The models are generated from the nonstationary parameters calibrated to the data (see Table 2 for the specific details of the model used to generate cumulative rainfall values).

added parameters for seasonality (seasonal covariates for occurrence and the marginal distribution); explicitly modeled event rainfall extremes in the marginal distribution; and the use of censored likelihood to coherently transform Gaussian fields into rainfall fields characterized by a large range of intermittency. Finally, the model features a new method of temporal disaggregation which uses the four-parameter beta distribution to relate maximum event intensity to cumulative event precipitation.

Stochastorm was applied to the Sahel region of West Africa. Specific adaptations included the categorization of small and large events, with the ratio evolving throughout the season (more large events during the peak precipitation months).

Results had good coherence with data, closely following the seasonal signal and annual properties of rainfall. The occurrence of rainfall events was well reproduced, although the current parameterization of the IET may not sufficiently simulate extreme values of annual number of events. The magnitude of events were remarkably reproduced including extreme values that are explicitly taken into account by the model. The

spatial intermittency and spatial structure of rain fields were also well replicated, in spite of the division into small and large events for the Sahel. The model calibrated and validated on two distinct sets of data produced comparable results, demonstrating the robustness of the model. However, as Stochastorm does not take long-term (decadal) temporal nonstationarity into account, simulations may deviate from the observations under the presence of nonstationarity in the rainfall series.

The use of a simple temporal disaggregation scheme based on a synthetic hyetograph maintained key rainfall characteristics in the simulations at subevent scales (from hourly to 5 min), although a slight underestimation of the 5-min rainfall variance has been noticed. The maximum intensity of the storms and their variability were better simulated via the stochastic beta distribution than the deterministic approach of the previous generator versions.

In addition to these noticeable performances, the explicit integration of seasonal variability in the parameters of occurrence and magnitude of events, the explicit representation of

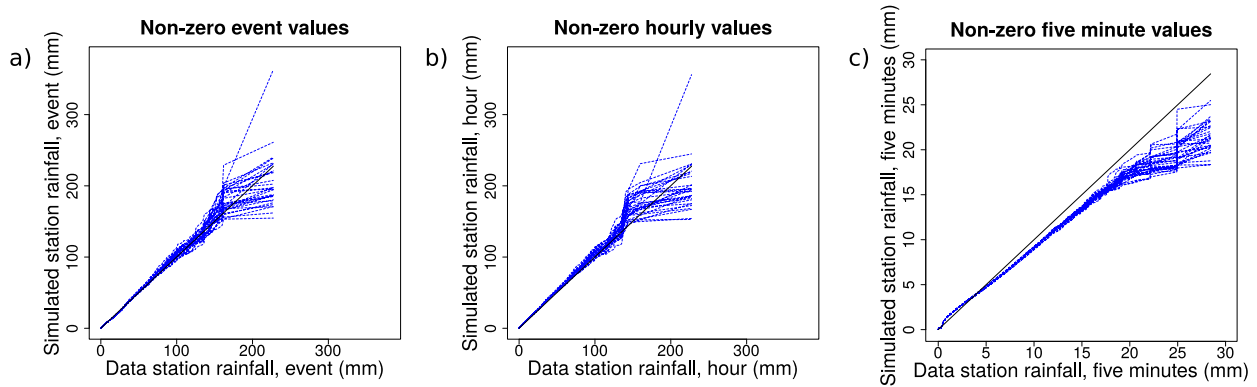


FIG. 10. Quantile–quantile plot for simulated vs observed rainfall by station at three temporal scales: event cumulative rainfall, hourly, and 5 min. There is one line per observation of 28 years or per simulation of 28 years (30 simulations total).

extremes in the marginal distribution and the possibility for simulating a wide range of intermittency make it an original contribution to recent advanced SRGs.

Stochastorm's parsimony, which is made possible by separating the simulation of cumulative events and their subsequent temporal disaggregation, is another notable asset in responding to simulation problems in regions where the available data (daily or hourly rainfall data) only provide access to storm accumulation structures but do not allow a precise description of subevent spatiotemporal structures. It is therefore a high-resolution SRG that has potential for applications in poorly instrumented developing countries, many of which are prone to tropical storms.

However, the simplified approach has some limits that should be addressed in future developments. In particular the case study pointed out the need to better simulate the variability of the number of annual events and 5-min rainfall amounts. Other assumptions such as the spatial stationarity of

the parameters can also be reconsidered to allow simulations on more complex terrains and/or rainfall typology than those of the Sahelian case study.

The simulation results have implications for applications to hydrological and agricultural modeling. Having simulation outputs at fine scales with a coherent and representative spatial structure is promising for implementation in hydrological models (Li et al. 2017). A priority for future work is testing the outputs of Stochastorm as inputs which drive a hydrological model (Wilcox 2019).

Stochastorm can also be calibrated on general circulation model (GCM) outputs, including future projections. The use of GCMs as calibration data would allow Stochastorm to function as a statistical downscaling tool (Ferraris et al. 2003) that can translate the GCM into relevant information at hydrological scales. This can provide a useful tool for evaluating potential future impacts of climate change.

The model is limited in its ability to handle long-term non-stationary, such as decadal trends. An example of ways to treat

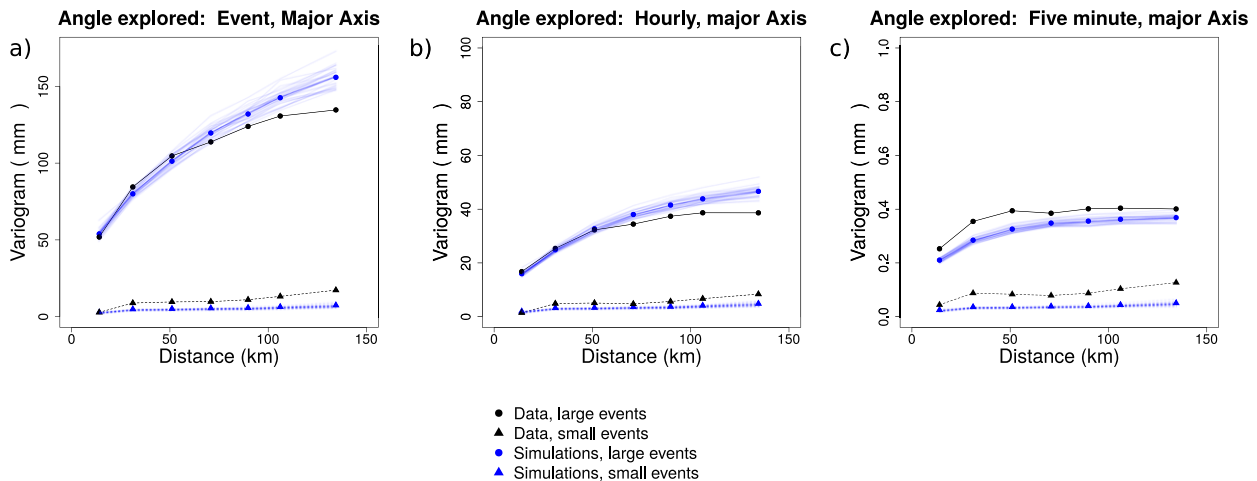
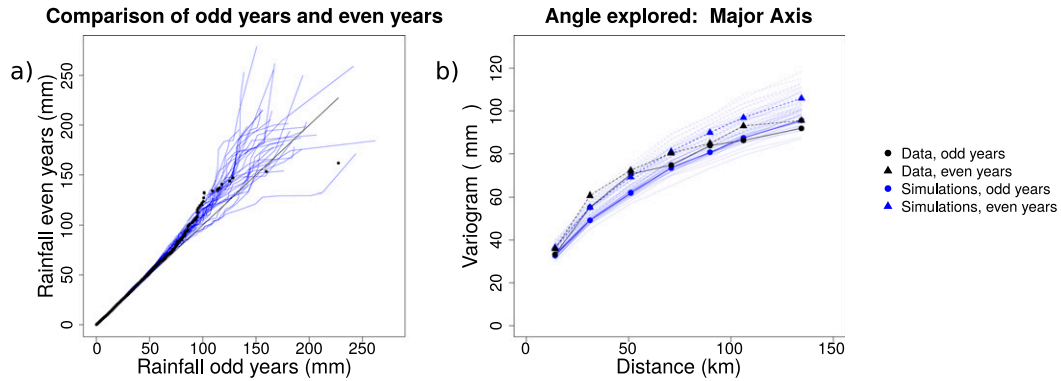


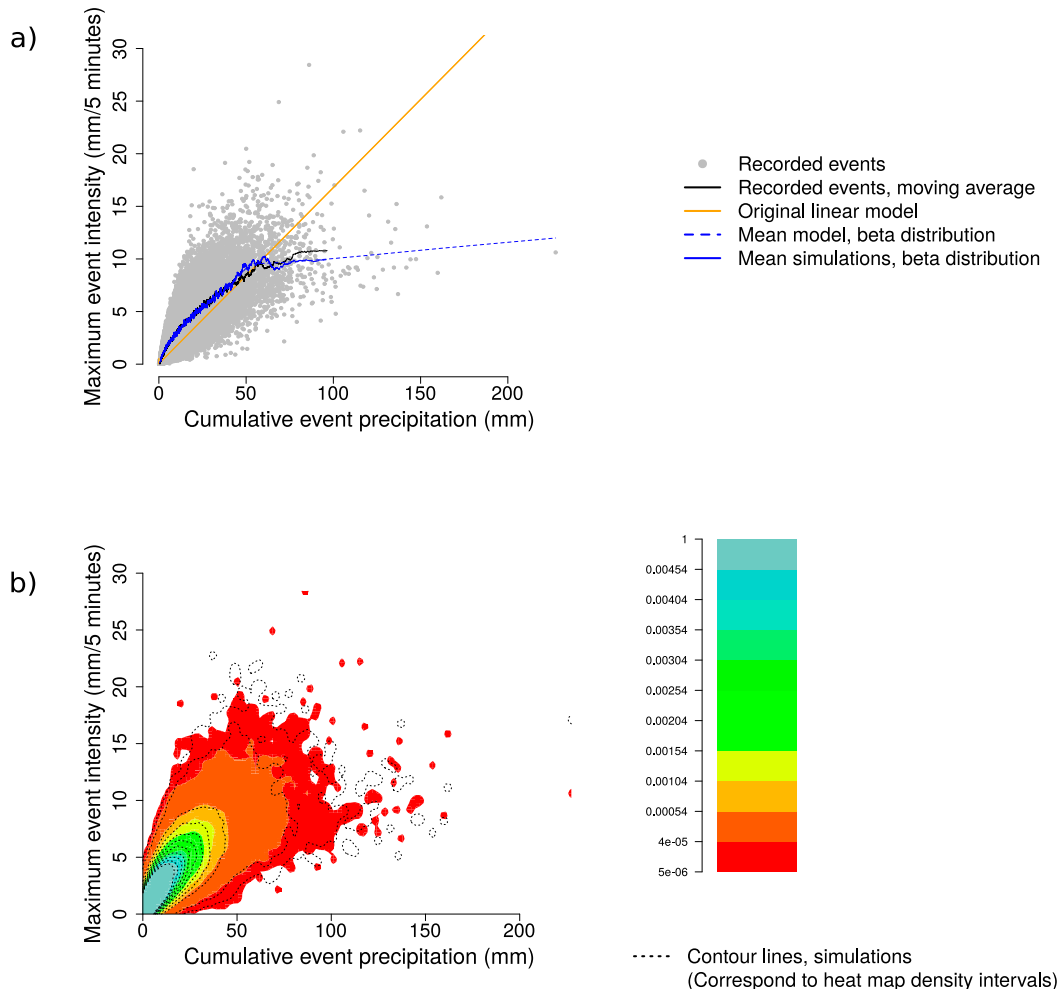
FIG. 11. Observed and simulated variograms for large and small events at three temporal scales: event cumulative rainfall, hourly, and 5 min. There is one line per observation of 28 years or per simulation of 28 years (30 simulations total). The points represent the average variogram for a given class of distances for all simulations.



nonstationarity with stochastic weather generators can be found in [Verdin et al. \(2018\)](#) and [Benoit et al. \(2018b\)](#). By incorporating nonstationary climate projections into the stochastic weather generator and coupling it with a hydrological

model, one gains the capacity to perform water resource projections ([Borgomeo et al. 2014](#)).

Although the model was implemented in the Sahel, it is based on precipitation characteristics that are common to



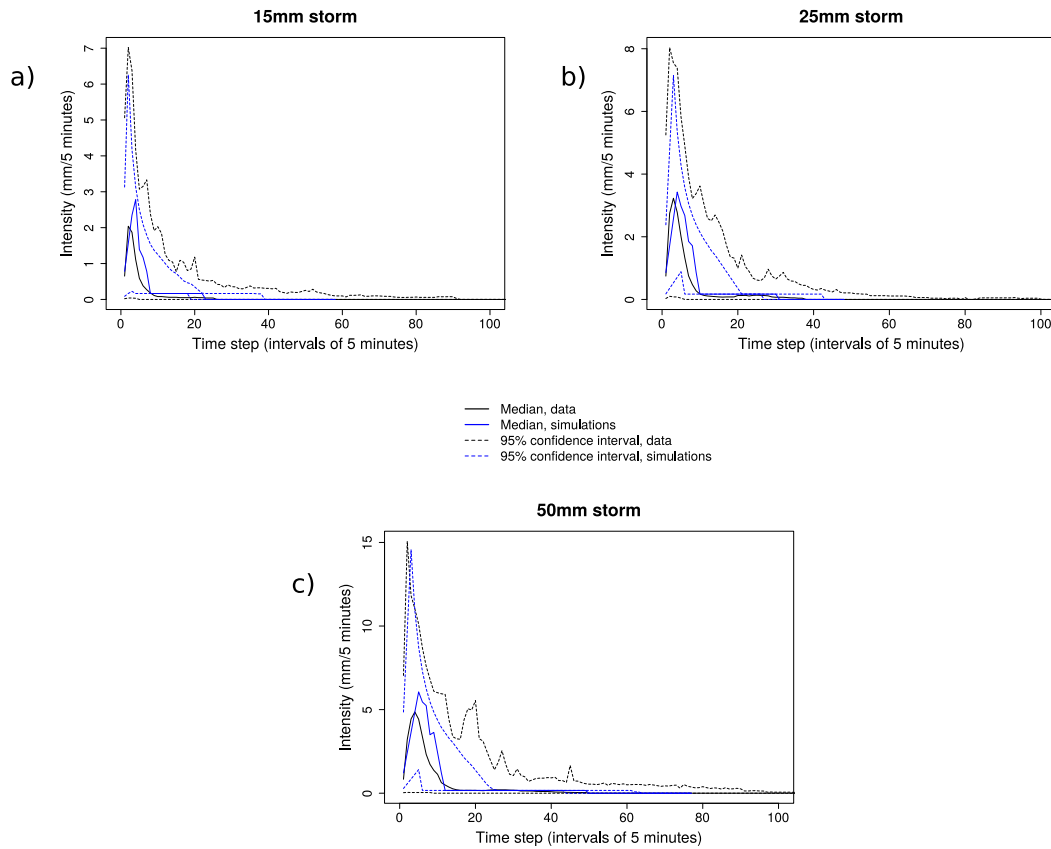


FIG. 14. Observed and simulated hyetographs corresponding to storms of different cumulative precipitation magnitudes (15, 25, and 50 mm).

convective storms in other regions of the world, especially semiarid ones. Stochastorm may be implementable in other regions provided that event-based parameters can be tuned on locally observed data and that an appropriate hyetograph shape and its link with event cumulative rainfall can be determined. This means that application to various study regions is not necessarily straightforward for Stochastorm. However, genericity is a problem that goes beyond the case of Stochastorm and concerns all SRGs. All SRGs are based on a given set of hypotheses, contexts, considerations, and data which would not be applicable in all parts of the world. Some future research directions would involve expanding the range of situations in which Stochastorm can perform well. There is also an interest to evaluate how other advanced high-resolution space-time SRGs as for instance those developed in Peleg et al. (2017) and Benoit et al. (2018a) would be able to adapt the Sahelian case study presented here. More generally, the question of genericity calls the community of high-resolution space-time SRGs to gather around intercomparison experiments based on various datasets from various areas.

Acknowledgments. The research leading to these results has received funding from the NERC/DFID Future Climate For Africa program under the AMMA-2050 project, Grant NE/M020428/1. The AMMA-CATCH regional observing

system (www.amma-catch.org) was set up thanks to an incentive funding of the French Ministry of Research that allowed pooling together various pre-existing small-scale observing setups. The continuity and long-term longevity of the measurements are made possible by undisrupted IRD funding since 1990 and by continuous CNRS-INSU funding since 2005. AMMA-CATCH also received support from OSUG, OREME, OMP, OSUG@2020 LabEx, SOERE RBV, and CRITEX EquipEx (Grant ANR-11- EOPX-0011).

Data availability statement. The raw precipitation data from the AMMA-CATCH observation network are available via the AMMA-CATCH database portal (<http://bd.amma-catch.org>). Their use is regulated by the Creative Commons Attribution 4.0 International License (CC-BY 4.0). The Stochastorm model presented in the article is available upon request in the form of an R package.

REFERENCES

- Ali, A., T. Lebel, and A. Amani, 2003: Invariance in the spatial structure of Sahelian rain fields at climatological scales. *J. Hydrometeorol.*, **4**, 996–1011, [https://doi.org/10.1175/1525-7541\(2003\)004<0996:IITSSO>2.0.CO;2](https://doi.org/10.1175/1525-7541(2003)004<0996:IITSSO>2.0.CO;2).
- AMMA-CATCH, 1990: Precipitation dataset (5 minutes rainfall), 30 long-term stations over the Niamey square degree site

- (16 000 km²), Niger. IRD, CNRS-INSU, OSUG, OMP, OREME, accessed 24 October 2018, <https://doi.org/10.17178/AMMA-CATCH.CL.Rain.N>.
- Arnaud, P., P. Cantet, and Y. Aubert, 2016: Relevance of an at-site flood frequency analysis method for extreme events based on stochastic simulation of hourly rainfall. *Hydrol. Sci. J.*, **61**, 36–49, <https://doi.org/10.1080/02626667.2014.965174>.
- Balme, M., T. Vischel, T. Lebel, C. Peugeot, and S. Galle, 2006: Assessing the water balance in the Sahel: Impact of small scale rainfall variability on runoff: Part I: Rainfall variability analysis. *J. Hydrol.*, **331**, 336–348, <https://doi.org/10.1016/j.jhydrol.2006.05.020>.
- Baxevani, A., and J. Lennartsson, 2015: A spatiotemporal precipitation generator based on a censored latent Gaussian field. *Water Resour. Res.*, **51**, 4338–4358, <https://doi.org/10.1002/2014WR016455>.
- Benoit, L., and G. Mariethoz, 2017: Generating synthetic rainfall with geostatistical simulations. *Wiley Interdiscip. Rev.: Water*, **4**, e1199, <https://doi.org/10.1002/war2.1199>.
- , D. Allard, and G. Mariethoz, 2018a: Stochastic rainfall modeling at sub-kilometer scale. *Water Resour. Res.*, **54**, 4108–4130, <https://doi.org/10.1029/2018WR022817>.
- , M. Vrac, and G. Mariethoz, 2018b: Dealing with non-stationarity in sub-daily stochastic rainfall models. *Hydrol. Earth Syst. Sci.*, **22**, 5919–5933, <https://doi.org/10.5194/hess-22-5919-2018>.
- Blanchet, J., C. Aly, T. Vischel, G. Panthou, Y. Sané, and M. D. Kane, 2018: Trend in the co-occurrence of extreme daily rainfall in West Africa since 1950. *J. Geophys. Res. Atmos.*, **123**, 1536–1551, <https://doi.org/10.1002/2017JD027219>.
- Borgomeo, E., J. W. Hall, F. Fung, G. Watts, K. Colquhoun, and C. Lambert, 2014: Risk-based water resources planning: Incorporating probabilistic nonstationary climate uncertainties. *Water Resour. Res.*, **50**, 6850–6873, <https://doi.org/10.1002/2014WR015558>.
- Breinl, K., G. Di Baldassarre, M. G. Lopez, M. Hagenlocher, G. Vico, and A. Rutgersson, 2017: Can weather generation capture precipitation patterns across different climates, spatial scales and under data scarcity? *Sci. Rep.*, **7**, 5449, <https://doi.org/10.1038/s41598-017-05822-y>.
- Burt, C. C., 2007: *Extreme Weather: A Guide & Record Book*. WW Norton & Company, 320 pp.
- Chardon, J., B. Hingray, A.-C. Favre, P. Autin, J. Gailhard, I. Zin, and C. Obled, 2014: Spatial similarity and transferability of analog dates for precipitation downscaling over France. *J. Climate*, **27**, 5056–5074, <https://doi.org/10.1175/JCLI-D-13-00464.1>.
- Cowpertwait, P., P. O'Connell, A. Metcalfe, and J. Mawdsley, 1996: Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. *J. Hydrol.*, **175**, 47–65, [https://doi.org/10.1016/S0022-1694\(96\)80005-9](https://doi.org/10.1016/S0022-1694(96)80005-9).
- D'Amato, N., and T. Lebel, 1998: On the characteristics of the rainfall events in the Sahel with a view to the analysis of climatic variability. *Int. J. Climatol.*, **18**, 955–974, [https://doi.org/10.1002/\(SICI\)1097-0088\(199807\)18:9<955::AID-JOC236>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0088(199807)18:9<955::AID-JOC236>3.0.CO;2-6).
- Descroix, L., P. Genthon, O. Amogu, J.-L. Rajot, D. Sighomnou, and M. Vauclin, 2012: Change in Sahelian rivers hydrograph: The case of recent red floods of the Niger River in the Niamey region. *Global Planet. Change*, **98–99**, 18–30, <https://doi.org/10.1016/j.gloplacha.2012.07.009>.
- Emery, X., 2002: Conditional simulation of nongaussian random functions. *Math. Geol.*, **34**, 79–100, <https://doi.org/10.1023/A:1014031528091>.
- Evin, G., A.-C. Favre, and B. Hingray, 2018: Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrol. Earth Syst. Sci.*, **22**, 655–672, <https://doi.org/10.5194/hess-22-655-2018>.
- Féral, L., H. Sauvageot, L. Castanet, and J. Lemorton, 2003: Hycell—A new hybrid model of the rain horizontal distribution for propagation studies: 1. Modeling of the rain cell. *Radio Sci.*, **38**, 1056, <https://doi.org/10.1029/2002RS002802>.
- Ferraris, L., S. Gabellani, N. Reborra, and A. Provenzale, 2003: A comparison of stochastic models for spatial rainfall downscaling. *Water Resour. Res.*, **39**, 1368, <https://doi.org/10.1029/2003WR002504>.
- Furrer, E. M., and R. W. Katz, 2008: Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resour. Res.*, **44**, W12439, <https://doi.org/10.1029/2008WR007316>.
- Guillot, G., 1999: Approximation of Sahelian rainfall fields with meta-Gaussian random functions. *Stochastic Environ. Res. Risk Assess.*, **13**, 100–112, <https://doi.org/10.1007/s004770050034>.
- , and T. Lebel, 1999a: Approximation of sahelian rainfall fields with meta-Gaussian random functions. *Stochastic Environ. Res. Risk Assess.*, **13**, 113–130, <https://doi.org/10.1007/s004770050035>.
- , and —, 1999b: Disaggregation of Sahelian mesoscale convective system rain fields: Further developments and validation. *J. Geophys. Res.*, **104**, 31 533–31 551, <https://doi.org/10.1029/1999JD900986>.
- Guilloteau, C., R. Roca, M. Gosset, and V. Venugopal, 2018: Stochastic generation of precipitation fraction at high resolution with a multiscale constraint from satellite observations. *Quart. J. Roy. Meteor. Soc.*, **144**, 176–190, <https://doi.org/10.1002/qj.3314>.
- Lebel, T., I. Braud, and J.-D. Creutin, 1998: A space-time rainfall disaggregation model adapted to Sahelian mesoscale convective complexes. *Water Resour. Res.*, **34**, 1711–1726, <https://doi.org/10.1029/98WR00434>.
- , A. Diedhiou, and H. Laurent, 2003: Seasonal cycle and interannual variability of the Sahelian rainfall at hydrological scales. *J. Geophys. Res.*, **108**, 8389, <https://doi.org/10.1029/2001JD001580>.
- Leblois, E., and J.-D. Creutin, 2013: Space-time simulation of intermittent rainfall with prescribed advection field: Adaptation of the turning band method. *Water Resour. Res.*, **49**, 3375–3387, <https://doi.org/10.1002/wrcr.20190>.
- Li, Z., Z. Lü, J. Li, and X. Shi, 2017: Links between the spatial structure of weather generator and hydrological modeling. *Theor. Appl. Climatol.*, **128**, 103–111, <https://doi.org/10.1007/s00704-015-1691-8>.
- Lombardo, F., E. Volpi, D. Koutsoyiannis, and F. Serinaldi, 2017: A theoretically consistent stochastic cascade for temporal disaggregation of intermittent rainfall. *Water Resour. Res.*, **53**, 4586–4605, <https://doi.org/10.1002/2017WR020529>.
- Mathon, V., H. Laurent, and T. Lebel, 2002: Mesoscale convective system rainfall in the Sahel. *J. Appl. Meteor.*, **41**, 1081–1092, [https://doi.org/10.1175/1520-0450\(2002\)041<1081:MCSRIT>2.0.CO;2](https://doi.org/10.1175/1520-0450(2002)041<1081:MCSRIT>2.0.CO;2).
- Onibon, H., T. Lebel, A. Afouda, and G. Guillot, 2004: Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resour. Res.*, **40**, W08401, <https://doi.org/10.1029/2003WR002009>.
- Panthou, G., T. Vischel, and T. Lebel, 2014: Recent trends in the regime of extreme rainfall in the Central Sahel. *Int. J. Climatol.*, **34**, 3998–4006, <https://doi.org/10.1002/joc.3984>.
- , and Coauthors, 2018: Rainfall intensification in tropical semi-arid regions: The Sahelian case. *Environ. Res. Lett.*, **13**, 064013, <https://doi.org/10.1088/1748-9326/AAC334>.

- Paschalis, A., P. Molnar, S. Fatichi, and P. Burlando, 2013: A stochastic model for high-resolution space-time precipitation simulation. *Water Resour. Res.*, **49**, 8400–8417, <https://doi.org/10.1002/2013WR014437>.
- Peleg, N., E. Shamir, G. Konstantine, and E. Morin, 2015: A framework for assessing hydrological regime sensitivity to climate change in a convective rainfall environment: A case study of two medium-sized eastern Mediterranean catchments, Israel. *Hydrol. Earth Syst. Sci.*, **19**, 567–581, <https://doi.org/10.5194/hess-19-567-2015>.
- , S. Fatichi, A. Paschalis, P. Molnar, and P. Burlando, 2017: An advanced stochastic weather generator for simulating 2-D high-resolution climate variables. *J. Adv. Model. Earth Syst.*, **9**, 1595–1627, <https://doi.org/10.1002/2016MS000854>.
- , C. Skinner, S. Fatichi, and P. Molnar, 2020: Temperature effects on the spatial structure of heavy rainfall modify catchment hydro-morphological response. *Earth Surf. Dyn.*, **8**, 17–36, <https://doi.org/10.5194/esurf-8-17-2020>.
- Peres, D., and A. Cancelliere, 2018: Modeling impacts of climate change on return period of landslide triggering. *J. Hydrol.*, **567**, 420–434, <https://doi.org/10.1016/j.jhydrol.2018.10.036>.
- Raut, B. A., A. W. Seed, M. J. Reeder, and C. Jakob, 2018: A multiplicative cascade model for high-resolution space-time downscaling of rainfall. *J. Geophys. Res. Atmos.*, **123**, 2050–2067, <https://doi.org/10.1002/2017JD027148>.
- Renard, B., D. Kavetski, E. Leblois, M. Thyer, G. Kuczera, and S. W. Franks, 2011: Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resour. Res.*, **47**, W11516, <https://doi.org/10.1029/2011WR010643>.
- Schleiss, M., S. Chamoun, and A. Berne, 2014: Stochastic simulation of intermittent rainfall using the concept of “dry drift”. *Water Resour. Res.*, **50**, 2329–2349, <https://doi.org/10.1002/2013WR014641>.
- Serinaldi, F., 2010: Multifractality, imperfect scaling and hydrological properties of rainfall time series simulated by continuous universal multifractal and discrete random cascade models. *Nonlinear Processes Geophys.*, **17**, 697–714, <https://doi.org/10.5194/npg-17-697-2010>.
- Sighomnou, D., and Coauthors, 2013: La crue de 2012 à Niamey: un paroxysme du paradoxe du Sahel? *Sci. Changements Planet./Secheresse*, **24**, 3–13.
- Sørup, H. J. D., O. B. Christensen, K. Arnbjerg-Nielsen, and P. S. Mikkelsen, 2016: Downscaling future precipitation extremes to urban hydrology scales using a spatio-temporal Neyman–Scott weather generator. *Hydrol. Earth Syst. Sci.*, **20**, 1387–1403, <https://doi.org/10.5194/hess-20-1387-2016>.
- Stehlik, J., and A. Bárdossy, 2002: Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. *J. Hydrol.*, **256**, 120–141, [https://doi.org/10.1016/S0022-1694\(01\)00529-7](https://doi.org/10.1016/S0022-1694(01)00529-7).
- Taylor, C. M., and Coauthors, 2017: Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. *Nature*, **544**, 475–478, <https://doi.org/10.1038/nature22069>.
- Troutman, B. M., 1983: Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resour. Res.*, **19**, 791–810, <https://doi.org/10.1029/WR019i003p00791>.
- Vaithinada Ayar, P., J. Blanchet, E. Paquet, and D. Penot, 2019: Space-time simulation of precipitation based on weather pattern sub-sampling and meta-Gaussian model. *J. Hydrol.*, **581**, 124451, <https://doi.org/10.1016/J.JHYDROL.2019.124451>.
- Verdin, A., B. Rajagopalan, W. Kleiber, G. Podestá, and F. Bert, 2018: A conditional stochastic weather generator for seasonal to multi-decadal simulations. *J. Hydrol.*, **556**, 835–846, <https://doi.org/10.1016/j.jhydrol.2015.12.036>.
- Vischel, T., T. Lebel, S. Massuel, and B. Cappelaere, 2009: Conditional simulation schemes of rain fields and their application to rainfall–runoff modeling studies in the Sahel. *J. Hydrol.*, **375**, 273–286, <https://doi.org/10.1016/j.jhydrol.2009.02.028>.
- Wilcox, C., 2019: Evaluating hydrological changes in semi-arid West Africa: Detection of past trends in extremes and framework for modeling the future. Ph.D. thesis, University of Joseph Fourier, 217 pp.
- , and Coauthors, 2018: Trends in hydrological extremes in the Senegal and Niger Rivers. *J. Hydrol.*, **566**, 531–545, <https://doi.org/10.1016/j.jhydrol.2018.07.063>.
- Wilks, D. S., 1998: Multisite generalization of a daily stochastic precipitation generation model. *J. Hydrol.*, **210**, 178–191, [https://doi.org/10.1016/S0022-1694\(98\)00186-3](https://doi.org/10.1016/S0022-1694(98)00186-3).
- , 2010: Use of stochastic weather generators for precipitation downscaling. *Wiley Interdiscip. Rev.: Climate Change*, **1**, 898–907, <https://doi.org/10.1002/wcc.85>.
- , and R. L. Wilby, 1999: The weather generation game: A review of stochastic weather models. *Prog. Phys. Geogr.*, **23**, 329–357, <https://doi.org/10.1177/030913339902300302>.
- Wilson, C. B., J. B. Valdes, and I. Rodriguez-Iturbe, 1979: On the influence of the spatial distribution of rainfall on storm runoff. *Water Resour. Res.*, **15**, 321–328, <https://doi.org/10.1029/WR015i002p00321>.