Statistical distribution of the fractional area affected by rain

Nicolas Jeannin, Laurent Féral, Henri Sauvageot, Laurent Castanet, and Joël Lemorton

Received 2 January 2008; revised 25 July 2008; accepted 2 September 2008; published 13 November 2008.

The knowledge of the fraction of an area that is affected by rain (or fractional area) is of prime interest for hydrologic studies or for rainfall field modeling. Up to now, the statistical distribution of this parameter has been poorly studied. In the present paper, a model of the statistical distribution of the fraction of an area affected by rain over a given rainfall rate is proposed. It takes into account at the same time the size of the area and the local climatology. The analytic formulation of the distribution is established, considering that rainfall fields can be obtained from a nonlinear filtering of a Gaussian random field. As the analytic derivation of the distribution lies on some assumptions, the model accuracy is first evaluated from numerical simulations. It is then shown that the model reproduces accurately the distribution of fractional areas derived from radar observations of rain fields for various rain thresholds, sizes of area, and climatologies. A generic parameterization is then proposed for areas ranging from 100 × 100 to 300 × 300 km².


1. Introduction

In a typical general circulation model (GCM), the spatial resolution of the outputs is generally not sufficient to describe rain at a scale of interest for hydrologic processes such as interception and runoff [Pitman, 1991; Thomas and Henderson-Sellers, 1991]. Therefore a representation of rainfall spatial variability at a finer scale is necessary. Thus many models were developed to assess the spatial variability of rain at coarser scales, using either fractal approaches [Lovejoy and Schertzer, 1985; Over and Gupta, 1996], stochastic approaches [Mejia and Rodriguez-Iurbe, 1974; Bell, 1987; Lebel et al., 1998] or cellular approaches [Le Cam, 1961; Capsoni et al., 1987a, 1987b; Goldhirsh 2000; Féral et al., 2003a, 2003b]. For most of these approaches, a key parameter that has to be specified is the fraction of the simulation area affected by rain, often called lacunarity or intermittency. Moreover, it was shown that surface hydrology exhibits a strong sensitivity to this parameter [Pitman et al., 1990]. Indeed, if the rain fractional coverage over an area of interest and for a given rain amount decreases, the climatology turns from evaporation dominated to runoff dominated, thus showing the importance of this quantity on climate simulations.

In other respects, the knowledge of this parameter is also valuable to evaluate the performances of earth-space telecommunication systems. Indeed, the attenuation undergone by an earth space satellite link operating at a frequency above 20 GHz is mainly driven by the rainfall rate along the path of the link [Castanet et al., 2001]. High attenuations due to rain will result in an unavailability of the satellite link or will require the use of adaptive fade mitigation techniques [Castanet et al., 2002; Neely et al., 2003]. Considering a satellite spot beam whose diameter is typically of 300 km, the knowledge of the fraction of the area affected by a rain rate over a given threshold gives an estimate of the fraction of the user-terminals in the satellite spot beam that will undergo a given level of attenuation. Consequently, this parameter gives an estimate, considering the link budget, either of the number of the terminals that will be in outage or of the additional margin that has to be provided to ensure a sufficient quality of service.

A convenient way to prescribe the fractional area f was addressed from the estimation of rainfall from space. Indeed, Donnœud et al. [1984], Chiu [1988], Braud et al. [1993], and Oki et al. [1997] observed that the rain fractional coverage over a preset threshold is tightly correlated with the spatially averaged rain rate over the area of interest. Particularly, whenever the local probability density function (PDF) of rain is known, Atlas et al. [1990], Kedem et al. [1990], and Sauvageot [1994] have shown that the proportionality coefficient between the fractional area and the spatially averaged rain rate can be determined from the rain conditional PDF (i.e., knowing that it is raining). Invoking an ergodicity assumption of the rainfall process, Eltahir and Bras [1993] gave a procedure to deduce the fractional area affected by rain from GCM outputs accounting for the GCM spatial and temporal resolutions. Nevertheless, for rain field simulation purposes, either as an input of the model [Féral et al., 2006] or as a tool to evaluate the model accuracy [Guillot, 1999; Guillot and Lebel, 1999], it...
could be interesting to have a simple way to get the statistical distribution of the rain fractional coverage for various thresholds, while accounting not only for the local climatology but also for the size of the simulation area.

[5] This was attempted by Onof and Wheater [1996] who tried, from weather radar observations over Wales, to approximate the PDF of the fractional area affected by rain with a parabolic distribution. In other respects, from radar observations over the French territory, Féral et al. [2006] approximated the distribution of the fractional area affected by rain with an exponential law with mean 8.8%. Nevertheless, in both studies, the dependence on the local climatology and the size of the area are not evaluated. Moreover, only the rain/no-rain threshold is considered, so that the results highly depend on the radar sensitivity.

[6] From an analysis of radar data on various climatic zones and for various sizes of observation area, this paper proposes an analytical expression of the cumulative distribution function (CDF) of the fractional area affected by rain over a preset threshold. The formulation is derived from properties of rain fields generated using a nonlinear transformation on stationary Gaussian random fields, a generation scheme that has already been reported in the literature [Bell, 1987; Guillot, 1999]. This approach has been theoretically justified by Ferraris et al. [2002]. The latter is based on two steps. First, a correlated, stationary, homogeneous Gaussian random field is generated on a lattice. Second, on each grid point, the Gaussian random variables are turned into rain rate values \( R \). More specifically, in the studies of Bell [1987] or Guillot [1999], the random values of the Gaussian field lower than a value \( \alpha \) driven by the local probability of rain are put to 0 while the random values greater than \( \alpha \) are transformed so that they follow the rain rate local conditional PDF \( p(R|R > 0) \). Nevertheless, as underlined by Guillot [1999], the main defect of such kind of modeling is that the fraction of area affected by rain on one realization is not prescribed. Furthermore, Guillot [1999] pointed out that if a long-range dependence is not introduced in the correlation function of the Gaussian random field, the fraction of the area affected by rain converges toward the local temporal expectancy to have rain, as a second-order stationary random field is ergodic. This means that the variance of the fractional coverage of rain computed from a large number of simulated fields tends to 0. However, and as shown by Eltahir and Bras [1993], the fractional coverage of rain is a highly variable parameter. Therefore rain fields simulated with the methodology described above must have a long-range dependence to be realistic. This long-range dependence can be quantified by what Lantuéjoul [1991] called the “integral range,” related to the average value of the correlation function over the domain of interest.

[7] Section 2 of the present paper precisely addresses the effect of the long-range dependence on the statistical properties of the spatial average and spatial variance for homogeneous stationary Gaussian fields generated on a finite grid. Particularly, it is shown that, under some assumptions, the spatial distribution of the samples defining one realization of a Gaussian field generated on a finite grid is approximately normal with mean \( M \) and variance \( 1 - \sigma^2 \), where \( \sigma^2 \) is the average value of the correlation matrix that drives the dependencies between all the points of the grid. Additionally, the spatial mean \( M \) is shown to be a random variable that follows a centered normal distribution with variance \( \sigma^2 \). From those considerations, in section 3, an analytical formulation is proposed to model the CDF of the fractional area of a stationary Gaussian random field over a preset threshold \( \alpha \). The model accuracy is first evaluated numerically from simulated Gaussian random fields. Then, in section 4, the model is compared with fractional area CDFs derived from true rain fields observed by weather radar over various places. A set of parameters is then proposed in order to reproduce the CDF of the fractional coverage of rain for various climatic zones and various sizes of area, ranging from \( 100 \times 100 \text{ km}^2 \) to \( 300 \times 300 \text{ km}^2 \).

2. Theoretical Considerations on Stationary Gaussian Random Fields

[5] As mentioned in the introductive part, the simulation of rain field using a nonlinear transformation of a Gaussian field has been already widely studied [Bell, 1987; Guillot, 1999] and was seen to perform satisfactorily in numerous situations. In those works, a methodology to derive the correlation function of the underlying Gaussian field from observed rain fields on a range of a few hundreds of kilometers was developed. The correlation function of the Gaussian field computed from rainfall observations (radar, rain gauges) was found to have a steep decay in the first tens of kilometers and was then only very slowly decreasing without reaching 0 at the end of the observation range of some hundreds of kilometers. Considering rain gauges spread across Italy, Barbalisica et al. [1992] and Bertorelli and Paraboni [2005] identified three scales of evolution for the correlation of rain fields: (1) below 100 km the correlation was shown to have a steep decay; (2) from 100 to 700 km the correlation was shown to be almost steady or very slowly decreasing with distance; and (3) after 700 km the correlation decreases slowly until total decorrelation.

[5] This long-range dependence can be explained by the non-uniform partition in space over wide areas of the rainy structures and of the underlying pressure and humidity field. This shape of correlation function is bothering because it implies that the rain fields are stationary over wide areas. As rain fields display some preferential path due for instance to the orography or to land sea transitions, this assumption would probably collapse for some hundreds of kilometers. The range of validity of this stationarity hypothesis is hardly assessed as the requirements to apply a formal stationarity test for spatial stochastic processes as described by Fuentes [2005] are not matched by rain fields. Consequently, the areas of side greater than \( 300 \times 300 \text{ km}^2 \) for which the stationary hypothesis of the rain fields is likely to be unrealistic will not be considered in the present paper.

[10] The aim of the rest of this section is to derive from those models an expression of the distribution of the fraction of an area \( f \) (or fractional area) over a given rain rate threshold using the properties of the underlying Gaussian field. The methodology developed in the following strives at finding an approximation of the spatial distribution of one realization of a Gaussian field on a finite grid considering that it has a correlation function of the shape described above. In section 3, from this spatial distribution, an expression of the fraction of the area of one realization of a Gaussian field is first derived. Then, the statistical distribution of the fractional area considering a large number of independent realizations
of the Gaussian field is proposed, validated and transposed to rain fields in section 4.

2.1. Spatial Average \( M \) of a Gaussian Random Field

[11] A stationary, homogeneous, standard, and centered Gaussian random field \( g \) generated on a bidimensional grid \( L \) with size \( N \times N \) can be interpreted as a set of standard centered Gaussian random variables \( g = (g(s_1), g(s_2), \ldots, g(s_{N^2})) \) correlated with each other, and indexed by their position \( s_i \) on the grid. Let \( \Sigma \) be the \( N^2 \times N^2 \) correlation matrix of the Gaussian random field \( g \) with general term \( \Sigma_{ij} = E[g(s_i)g(s_j)] \), where \( E[\cdot] \) is the expectancy operator. The spatial average \( M \) of the random field \( g \) over the grid \( L \) is defined by:

\[
M = \frac{1}{N^2} \sum_{i=1}^{N^2} g(s_i). \tag{1}
\]

As all the random variables \( g(s_i) \) follow a standard centered normal distribution, the random variable \( M \) is also normal with an expectation of 0. The variance of \( M \) is defined by:

\[
\text{Var}(M) = \frac{1}{N^2} E \left[ \left( \sum_{i=1}^{N^2} g(s_i) \right)^2 \right] = \frac{1}{N^2} E \left[ \sum_{i=1}^{N^2} g(s_i)g(s_j) \right] = \frac{1}{N^2} \sum_{i=1}^{N^2} \sum_{j=1}^{N^2} \Sigma_{ij} = \sigma^2. \tag{2}
\]

[12] Equation (2) shows that the variance \( \text{Var}(M) \) of the spatial average \( M \) is the mean \( \sigma^2 \) of the terms of the correlation matrix \( \Sigma \). At this stage we demonstrate that the spatial average \( M \) of the field is normally distributed with mean 0 and variance \( \sigma^2 \).

2.2. Spatial Variance \( V \) of a Gaussian Random Field

[13] Similarly, the spatial variance \( V \) of the Gaussian random field \( g \) over the grid is given by:

\[
V = \frac{1}{N^2} \sum_{i=1}^{N^2} (g(s_i) - M)^2. \tag{3}
\]

The spatial variance \( V \) is a random variable and its expectation can be defined as:

\[
E[V] = \frac{1}{N^2} \sum_{i=1}^{N^2} E \left[ (g(s_i) - M)^2 \right] = \frac{1}{N^2} \sum_{i=1}^{N^2} E \left[ g(s_i)^2 \right] - 2E[g(s_i)M] + E[M^2] = \frac{1}{N^2} \sum_{i=1}^{N^2} \left\{ \text{Var}[g(s_i)] + \text{Var}[M] - \frac{2}{N^2} \sum_{j=1}^{N^2} E[g(s_i)g(s_j)] \right\} = 1 + \text{Var}(M) - \frac{1}{N^4} \sum_{i=1}^{N^2} \sum_{j=1}^{N^2} \Sigma_{ij} = 1 - \text{Var}(M) = 1 - \sigma^2. \tag{4}
\]

[14] Consequently, the samples \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) of one realization \( g(\omega) \) of the Gaussian random field \( g \) are characterized by their spatial mean \( M(\omega) \) and by their spatial variance \( V \). In compliance with section 2.1, the spatial mean \( M(\omega) \) is drawn from a normal distribution whose mean is 0 and whose variance \( \sigma^2 \) is the mean of the correlation function \( \Sigma \). In such conditions, the expectancy \( E[V] \) of the spatial variance \( V \) is \( 1 - \sigma^2 \), as shown by equation (4).

The random field \( g \) can thus be considered as the weighted sum of a standard centered Gaussian random field \( \tilde{g} \) whose correlation is \( c_\tilde{g} \) and weight \( b \) with a standard centered Gaussian random variable \( y \) and weight \( a \). Consequently, for each point \( s \) of the grid:

\[
g(s) = bg(s) + ay. \tag{6}
\]

If \( c_\tilde{g} \) rapidly decreases with regards to the size of the lattice, the spatial distribution of \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) tends to be normal. Moreover, as \( y \) does not vary in space, the spatial distribution of \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) also tend to be normal. Considering the correlation functions proposed for rain field simulation by Barbaliscia et al. [1992] or Guillot and Lebel [1999], and grid sizes ranging from \( 100 \times 100 \) to \( 300 \times 300 \) km\(^2\), the steep decay \( c_\tilde{g} \) with regards to the size of the grid is ensured and the normality of \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) is a reasonable assumption. As no means were found to quantify the derivation from normality of \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) due to the finiteness of the lattice, the effect of this approximation on the accuracy of the model proposed in section 3.1 is assessed from simulated Gaussian fields in section 3.2.

3. Model for the CDF of the Fractional Area Over a Given Threshold

3.1. Analytical Derivation

[15] As discussed in section 2.2, for one realization \( g(\omega) \) of a stationary standard centered Gaussian random field \( g \) with suitable shape of correlation function, the spatial distribution of \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) tends to be normal if the size of the grid \( L \) is sufficiently large. For the purpose of this study, the spatial distribution \( \{g(s_i, \omega)\}_{i=1, \ldots, N^2} \) is
considered as a normal distribution with mean $M$ and variance $V = 1 - \sigma^2$. Besides, as shown in section 2.1, $M$ is random and follows a centered normal distribution with variance $\sigma^2$. Therefore considering one realization $\omega$ of the homogeneous random field $g$, the fractional area $f_\alpha(\omega)$ of $g$ over which the threshold $\alpha$ is exceeded is given by:

$$f_\alpha(\omega) = \frac{1}{2} \text{erfc} \left( \frac{\alpha - M(\omega)}{\sqrt{2(1 - \sigma^2)}} \right).$$

(7)

where $\text{erfc}$ denotes the complementary error function.

[16] The CDF of the fractional area $f$ exceeding the threshold $\alpha$ can then be derived considering the CDF of $M$:

$$P_\alpha(f > f^*) = P \left( \frac{1}{2} \text{erfc} \left[ \frac{\alpha - M}{\sqrt{2(1 - \sigma^2)}} \right] > f^* \right)$$

$$= P \left( M > \alpha - \frac{1}{\sqrt{2(1 - \sigma^2)}} \text{erfc}^{-1}(2f^*) \right) .$$

(8)

Recalling that $M$ follows a centered normal law with variance $\sigma^2$, we have:

$$P(M > x) = \frac{1}{2} \text{erfc} \left( \frac{x}{\sqrt{2}\sigma} \right).$$

(9)

so that:

$$P_\alpha(f > f^*) = \frac{1}{2} \text{erfc} \left[ \frac{\alpha - \sqrt{2(1 - \sigma^2)} \text{erfc}^{-1}(2f^*)}{\sqrt{2}\sigma} \right].$$

(10)

Derivation of equation (10) gives the PDF $p_\alpha(f)$ of the fractional area $f$ above $\alpha$:

$$p_\alpha(f) = \exp \left\{ - \left[ \text{erfc}^{-1}(2f) \right]^2 \right\} \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ - \frac{\alpha - \text{erfc}^{-1}(2f)}{\sqrt{2(1 - \sigma^2)}} \right\} .$$

(11)

The analytical formulation of the fractional area average value $\langle f \rangle_\alpha$ is difficult to obtain from equation (11). Nevertheless, it can be accurately approximated by:

$$\langle f \rangle_\alpha = P(G > \alpha) = \frac{1}{2} \text{erfc} \left( \frac{\alpha}{\sqrt{2}} \right).$$

(12)

equation (12) is obtained by considering that, over several realizations of the Gaussian random field $g$, the fraction $f$ of field $g$ above the threshold value $\alpha$ is simply the overall number of points above $\alpha$ divided by the overall number of points. As $g$ is homogeneous, the expectancy to exceed $\alpha$ is the same everywhere in the lattice so that $\langle f \rangle_\alpha$ amounts to equation (12), i.e., to the probability for a Gaussian centred standard variable $G$ to exceed $\alpha$. The numerical computation of $\langle f \rangle_\alpha$ from the PDF (equation (11)) with different parameterizations complies satisfactorily with equation (12), i.e., with a maximum error lower than 1%.

[17] In compliance with equations (10) and (11), the statistical distribution of the fractional area $f$ depends on 2 parameters. The first one, $\alpha$, is related to the average value of $f$ by equation (12). The second one, $\sigma^2$, is the average value of the $N^2$ Gaussian field correlation matrix $\Sigma$. Obviously, $\sigma^2$ strongly depends on the size $N \times N$ of the simulation grid. So as to quantitatively assess the model accuracy and the parameter sensitivity, equation (10) is compared to the fractional area CDFs derived from Gaussian field simulation.

3.2. Model Accuracy and Parameter Sensitivity

[18] For each configuration detailed hereafter, 6000 stationary Gaussian fields are generated using the spectral method described by Bel [1987]. Further details on this approach can be found in the studies of Shinozuka and Jan [1972], Mejia and Rodriguez-Iturbe [1974], and Borbora et al. [1984]. Lattice size varies from $N \times N = 50 \times 50$ km$^2$ to $N \times N = 200 \times 200$ km$^2$, in compliance with the typical coverage of the operational radars considered in section 4. Moreover, to evaluate the model sensitivity, two analytical formulations of the correlation function are considered for the Gaussian field $g$: $c_g(d) = e^{-d/30}$ and $c_g(d) = 0.5 e^{-d/30} + 0.5 e^{-d/800}$, where $d$ is the distance in kilometers. Such exponential formulations of the correlation function are commonly accepted to simulate rain fields [Guillot and Lebel, 1999; Barbaliscia et al., 1992; Féral et al., 2006]. In such conditions, considering $c_g(d)$ and equation (2), $\sigma$ varies from 0.71 to 0.34 when $N \times N$ varies from 50 $\times$ 50 km$^2$ to $N \times N = 200 \times 200$ km$^2$, respectively. Similarly, considering $c_g(d)$, $\sigma$ varies from 0.85 to 0.72 when $N \times N$ varies from 50 $\times$ 50 km$^2$ to $N \times N = 200 \times 200$ km$^2$, respectively. Six threshold values $\alpha$ are considered successively to compute the fractional area CDFs from the simu-
lated Gaussian fields, namely $\alpha = 0.5, 1, 1.5, 2, 2.5, 3$. The distributions thus obtained are compared with equation (10). The results are shown for the correlation function $c_{g,2}(d)$ in Figures 1 and 2. It then appears that, whatever the configuration, model (10) accurately reproduces the fractional area CDFs derived from the simulated fields. As all the realization of the simulated field are independent the values $f$ obtained for each realization of the fields are independent and common statistical test can therefore be applied to test the validity of equation (10). Particularly, the null hypothesis “the CDF $P_{\alpha}(f > f^*)$ derived from the simulated Gaussian fields is equation (10)” is never rejected by a unilateral Smirnov-Kolmogorov test [Chakravarti et al., 1967] with a confidence level of 0.05. To conclude, model (10) accurately reproduces the Gaussian field fractional area CDFs whatever the threshold $\alpha$, the lattice size between $50 \times 50$ and $200 \times 200$ km$^2$, and the standard formulations $c_{g,1}(d)$ or $c_{g,2}(d)$ of the correlation function.

4. Application to Rain Fields

4.1. Model Extension to Rain Field

[19] In the previous section, a mathematical framework to describe the CDF of the fractional area of a homogeneous Gaussian field that exceeds a preset threshold $\alpha$ has been developed. The same methodology applies to rain fields whenever the latter are derived from a nonlinear transformation of a stationary homogeneous Gaussian field [Bell, 1987; Guillot, 1999]. Indeed, considering that the rain rate $R$ conditional PDF is lognormal (mean $\mu_R$, standard deviation $\sigma_R$) as it is commonly accepted [Bell, 1987; Sauvageot, 1994; Féral et al., 2006], and assuming that it rains a fraction $P_0$ of the time, the transformation $\psi$ that converts the Gaussian field $g$ into a rain field $r$ is:

$$r(s) = \psi[g(s)] = \begin{cases} 
0 & \text{if } g(s) < \alpha_{P_0} \\
\exp[\mu_R + \sqrt{2}\sigma_R \text{erfc}^{-1}\left(\frac{\text{erfc}(g(s)/\sqrt{2})}{P_0}\right)] & \text{if } g(s) \geq \alpha_{P_0},
\end{cases}$$

where

$$\alpha_{P_0} = \sqrt{2}\text{erfc}^{-1}(2P_0).$$

By construction, the expectation for a Gaussian value to be converted into rain is exactly $P_0$. Moreover, the computation of the CDF of the fraction $f$ of a rain field affected by a rain rate $R$ exceeding the threshold $R^*$ amounts to the computation of the fraction of a standard Gaussian field exceeding the threshold $\alpha_{R^*} = \psi^{-1}(R^*)$. As $P(r(s) > R^*) = P[g(s) > \alpha_{R^*}]$, we have:

$$\alpha_{R^*} = \sqrt{2}\text{erfc}^{-1}\left(2P[r(s) > R^*]\right).$$

If we assume that the rain field $r(s)$ is homogeneous, the probability of rain $P_0$ is the same all over the grid and $P[r(s) > R^*]$ does not depend on the location $s$ so that equation (15) reduces to:

$$\alpha_{R^*} = \sqrt{2}\text{erfc}^{-1}\left(2P[R > R^*]\right).$$

4.2. Radar Data

[20] Two yearly radar data sets are considered. The first one was collected in 1996 and comes from the weather radar of Bordeaux that is part of the French operational radar

![CDF of the fraction of the field over $\alpha$](image)

Figure 2. CDFs of the fractional area over the threshold $\alpha$ computed from model (11) (markers) and derived from the simulation of Gaussian fields (dashed line): size $N \times N = 50 \times 50$ km$^2$, correlation function $c_0(d) = 0.5\exp(-d/30) + 0.5\exp(-d/800)$.

where $P(R > R^*)$ is the rain absolute CDF intrinsic to the simulation area. Equation (16) gives a convenient way to determine $\alpha_R$ from radar data. Finally, the CDF of the fractional area $f$ of a rain field affected by a rain rate exceeding the threshold value $R^*$ derives from equation (10):

$$P_{R^*}(f > f^*) = \frac{1}{2}\text{erfc}\left[\frac{\alpha_{R^*} - \sqrt{2(1 - \sigma^2)}\text{erfc}^{-1}(2f^*)}{\sqrt{2}\sigma}\right],$$

under the hypothesis made to establish equations (10) and (17), namely the homogeneity of the field (i.e., the pdf of rainfall rate does not vary inside the considered area) and a correlation function with a fast and a slow decay component with respect to the lattice size as advocated by Barbalisca et al. [1992] or Guillot and Lebel [1999].
network managed by Météo-France. The second one refers to 2004 and consists of composite radar images from the US NEXRAD network. For both data sets, the temporal sampling is one observation every 5 min. The data were projected on a Cartesian grid within a uniform pixel size of $1 \times 1$ km$^2$. Images including ground clutter or melting layer echoes were removed from the data set so that the used radar data only refer to rainfall fields. Lastly, reflectivity fields were converted into rain fields using the standard Z-R relation:

$$Z = aR^b,$$

where $a = 300$, $b = 1.35$, and $Z$ is the radar reflectivity factor in mm$^6$ m$^{-1}$ and $R$ the rain rate in mm h$^{-1}$.

As shown in Figures 3 and 4, areas of interest offering clearly distinct climatologies have been selected inside the radar coverage areas. Moreover, for each selected location, five area sizes are considered successively, namely $100 \times 100$, $150 \times 150$, $200 \times 200$, $250 \times 250$ and $300 \times 300$ km$^2$.

4.3. Results

The CDFs of the fractional area affected by rain were computed for thresholds $R^* = 0.5, 1, 2.4, 5.7, 13$ mm h$^{-1}$ for the areas of interest over the United States and $R^* = 1, 1.4, 2.2, 3.5$ mm h$^{-1}$ for the ones over Bordeaux (the thresholds considered are not exactly the same for both data sets due to some differences in the quantization steps between the radar of Bordeaux and the NEXRAD network).

First, the rain rate absolute climatological probabilities $P(R > R^*)$ are computed by counting the number of pixels over $R^*$ on the whole data set and by dividing it by the overall number of pixels. These distributions may encounter more or less large fluctuations on the same geographical zone for the different area sizes. This may be due to some defects in radar measurements (particularly attenuation with increasing range) or to some climatologic inhomogeneities of rain processes in squares of some hundreds of kilometers of side [Onof and Wheater, 1996].

Then, for the successive values of $R^*$, the corresponding $\alpha_R$ is deduced from equation (16). The results are given in Table 1. For $R^* = 1$ mm h$^{-1}$ only, model (17) $F(f; \alpha_1, \sigma) = \frac{1}{2}\text{erfc} \left( \frac{\alpha_1 - \sqrt{2(1-\sigma^2)\text{erfc}^{-1}(2\sigma)}}{\sqrt{2\sigma}} \right)$ is regressed with respect to the fractional area CDF $P(f > f^*)$ derived from radar observations to obtain parameters $\sigma$ reported in Table 3 (first line, area size $N \times N = 100 \times 100$ km$^2$ located over the United States and over Bordeaux). The values of $\sigma$ were

![Figure 3. Geographical position of the three areas of interest selected over the United States. The background image is a rain field observation collected by the NEXRAD radar network on 27 August 2004 at 17:00 UTC.](image-url)
obtained solving equation (19) using a nonlinear least squares algorithm:

\[
\arg \min_{\alpha_1} \sum_i \left[ \frac{F(f_i, \alpha_1, \sigma) - P(R^*(f > f_i))}{P_R(f > f_i)} \right]^2,
\]

where \( \alpha_1 \) is determined from equation (16) and the \( f_i \) are regularly sampled from 0.01 to \( f_{\text{max}} \) by step of 0.01. \( f_{\text{max}} \) is defined as the 30th largest value of fractional area found on each radar data set for each considered threshold \( R^* \). This choice of \( f_{\text{max}} \) is made in order to ensure a sufficient statistical reliability to the empirical CDFs especially for the greatest values of occupation. This regression procedure allows defining a relative error criterion between the model (17) and the CDF regressed from radar data:

\[
\varepsilon = 100 \times \frac{1}{N} \sqrt{\sum \left[ \frac{F(f_i, \alpha_{R^*}, \sigma) - P_{R^*}(f > f_i)}{P_{R^*}(f > f_i)} \right]^2}.
\]

**Figure 4.** Geographical position of the area of interest over Bordeaux (France). The background image is an example of rain field observation collected by the ARAMIS radar network on 9 December 2001 at 16:00 UTC.
Equation (20) quantifies the error made when approximating the empirical occupation CDF by model (19) in terms of mean relative RMS error.

\[ s, \text{ obtained using this regression procedure for each data set are used for all the other values of } R^* \text{ as } s \text{ should not depend on the threshold whenever rain fields are stationary and homogeneous. Besides, } R^* = 1 \text{ mm h}^{-1} \text{ is the value considered in the regression process because, first, it is the lowest value common to both data sets, and second, it complies with Krajewski et al. [1992] that advocates the use of a relatively low threshold to better estimate the fractional area from real data.} \]

For the successive threshold values \( R^* \), Figures 5 and 6 show the fractional area CDFs derived from model (17) and those derived from the radar observations over Bordeaux.
(France), for area sizes $N \times N = 100 \times 100$ km$^2$ 

(Ohio) for area sizes $N \times N = 100 \times 100$ km$^2$ ($\sigma$ regressed for $R^* = 1$ mm h$^{-1}$ is 0.88) and $N \times N = 200 \times 200$ km$^2$ ($\sigma$ regressed for $R^* = 1$ mm h$^{-1}$ is 0.77), respectively. Similarly, the results obtained for Ohio (USA) are shown in Figure 7, for $N \times N = 100 \times 100$ km$^2$ ($\sigma$ regressed for $R^* = 1$ mm h$^{-1}$ is 0.9). Lastly, Figures 8 and 9 underline the fractional area CDF dependence on the geographical location and the size of the area, respectively. The impact of the location on the distribution is mainly driven by the local rain rate CDF while increasing the area size lowers the average value $\sigma^2$ of the correlation function.

Nevertheless, in compliance with Figures 5 to 9, model (17) satisfactorily compares with the fractional area CDFs derived from radar observations whatever the threshold $R^*$, the location, or the area size $N \times N$ ranging 100 to 300 km$^2$. However, some trends can be noticed considering the errors computed according to equation (20) for different areas, rain rate thresholds and locations as shown in Table 2. Firstly, the larger are the considered areas, the higher the errors between the two distributions. Secondly, for threshold values $R^* \geq 5$ mm h$^{-1}$ for US radar and $R^* \geq 3.4$ mm h$^{-1}$ for Bordeaux, the model tends to overestimate the probability of having a large fraction of the area affected by rain.

These two trends question the rain field homogeneity hypothesis used to derive equations (10) and (17). Indeed,
Table 3. Regressed Value of $\sigma$ for 2 Rain Rate Thresholds $R^*$ and for a $100 \times 100$ km$^2$ Area of Interest

<table>
<thead>
<tr>
<th></th>
<th>Dakota</th>
<th>Mississippi</th>
<th>Ohio</th>
<th>Bordeaux</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^* = 1$ mm h$^{-1}$</td>
<td>0.86</td>
<td>0.90</td>
<td>0.92</td>
<td>0.88</td>
</tr>
<tr>
<td>$R^* = 5$ mm h$^{-1}$</td>
<td>0.83</td>
<td>0.87</td>
<td>0.90</td>
<td>0.85 ($R^* = 3.4$ mm h$^{-1}$)</td>
</tr>
</tbody>
</table>

over areas of hundreds of square kilometers, the rain rate local CDF may differ from one point to the other. In such conditions, the model no longer holds because the stationarity and homogeneity of the field become an unrealistic assumption. In other respects, the differences observed for increasing values of $R^*$ may be linked to the use of a constant parameter to describe the average value $\sigma^2$ of the correlation function. Indeed, high rain rates result from a convective process. Now the radar observation of convective rain fields shows clustered structures, whose spatial extent is clearly lower than that observed for stratiform rain events. This point is partially confirmed when regressing model (17) for $R^* = 5$ mm h$^{-1}$ (US radar) and $R^* = 3.4$ mm h$^{-1}$ (Bordeaux) to determine $\sigma$. Indeed, the results reported in Table 3 show that, whatever the location, $\sigma$ is slightly lower than the value obtained for $R^* = 1$ mm h$^{-1}$, suggesting (as expected) that high rain rates are less correlated in space than lower rain rates because of the average of the correlation of the underlying Gaussian fields for rain fields thresholded with high values is lower than the one found for lower thresholds. This may be due to a faster decay of the correlation considering rain fields with a more significant proportion of convective rain.

4.4. Parameterization of the Distribution

The results obtained for the parameter $\sigma$ of the distribution display an interesting stability as shown in Figure 10.

Indeed, for different climatic regions, this parameter seems to have approximately the same behavior with regards to the evolution of the size of the area. A generic parameterization of $\sigma$ independent from the location can hence be regressed and given as an indication for areas ranging from $100 \times 100$ to $300 \times 300$ km$^2$:

$$\sigma = 0.94 - 0.0007L,$$

(21)

where $L$ is the length of the side in kilometers. The relatively high values found for $\sigma$ argue in favor of the existence of a long-range correlation confirming the hypothesis made in the second section of this study. Equation (21) allows to get for an arbitrary location an approximation of the distribution of the fraction of an area $L \times L$ affected by a rain rate over a given threshold, whenever the local distribution of rain rates is known. Indeed, from $\mu_R, \sigma^2_R$ and $P_o$, where $\mu_R, \sigma^2_R$ are the parameters of the lognormal pdf that characterizes the conditional distribution of rain rates and $P_o$ is the probability of rain, the parameter $\alpha$ needed to compute the distribution of the fraction of the area affected by rain can be deduced from equation (16).

When combined with equation (17), parameters $\alpha$ and $\sigma$ obtained that way allow to approximate the CDF of the fraction of an area affected by rain for areas ranging from $100 \times 100$ to $300 \times 300$ km$^2$. The lower bound of this domain is driven by the requirements on the size of the area with regards to the alleged underlying correlation function. The upper bound is set considering the hypothesis of stationarity of the rain fields and of spatial invariance of the point rainfall rate distribution across the areas that were considered to establish the model. It should be noticed that the data used in this study have been projected on a grid of $1 \times 1$ km$^2$. Considering data with a lower spatial resolution, the parameterization proposed will no longer hold. First, the standard deviation of the distribution of rainfall rate that is used to determine the thresholds will change. Second, the parameter $\sigma^2$ that corresponds to the spatial average of the correlation is likely to be slightly higher as the reduction of the resolution will smooth the data and consequently generates higher correlation.

5. Concluding Remarks and Applications

From considerations on stationary Gaussian random fields, a model that reproduces the CDF of the fractional area affected by rain over a preset threshold has been proposed for areas ranging from $100 \times 100$ to $300 \times 300$ km$^2$. It has been developed assuming that rain fields can be seen as a nonlinear transformation of a stationary Gaussian random field.

First, the model was shown to match very accurately the fractional area CDF derived from simulated Gaussian random fields. Second, when confronted to distributions obtained from radar observations of rain field, the model satisfactorily reproduces the CDF of the fractional area above a rain threshold $R^*$. This model requires two inputs. The first one $\alpha$ is related to the local distribution of rain rates, the second one $\sigma$ to the mean of the correlation function of the underlying Gaussian field. A parameterization accounting for the local climatology and for the size of the area was proposed for this parameter. The optimal value of $\sigma$ was found to be slightly dependent on the threshold considered on radar data, questioning the hypothesis of the stationarity of the field.

The high values found in any case for the parameter $\sigma$ put forward the long-range dependence that has to be
introduced in the correlation function of the Gaussian field when generating rainfall fields. Important conclusions can be drawn from those results for simulation purposes. For models such as those described by Bell [1987] or Guillot and Belbel [1999], the use of a correlation function that has a mean value of $\sigma^2$ allows, when generating a large collection of simulated fields, to reproduce the local distribution of rain rates and the distribution of the fraction of the area affected by rain over a given threshold. Additionally, the distribution of the fraction of the area affected by rain is required by rainfall field generators such as the one described by Féraud et al. [2006]. The methodology presented here provides a convenient mean to approximate this distribution without computing it from radar data.

[34] Acknowledgments. The authors are very grateful to Me.


