Inter-annual variability, risk and confidence intervals associated with propagation statistics. Part I: theory of estimation

Nicolas Jeannin¹, Xavier Boulanger¹,²,*†, Laurent Féral³, Laurent Castanet¹ and Frédéric Lacoste²

¹Electromagnetics and Radar Department, Radio-Communications and Propagation Research Unit, ONERA, 2 Avenue E. Belin, BP 74025, 31055, Toulouse Cedex 4, France
²DCT/RF/ITP, Centre National d’Etudes Spatiales (CNES), 18 Avenue E. Belin, 31401, Toulouse Cedex 4, France
³Groupe de Recherche en Electromagnétisme (GRE), Université Paul Sabatier - Laboratoire LAPLACE, 118 route de Narbonne, 31062, Toulouse Cedex 9, France

SUMMARY

This set of two companion papers aims at providing a statistical framework to quantify the inter-annual variability observed on the statistics of rain attenuation or rainfall rate derived from Earth-space propagation measurements. This part I is more specifically devoted to the theoretical study of the variance of estimation of empirical complementary cumulative distribution functions (ECCDFs) derived from Earth-space rain attenuation or rainfall rate time series. To focus the analysis on the statistical variability but without loss of generality, synthetic rain attenuation time series are considered. A large variability on the ECCDFs, which depends on the duration of the synthetic data, is first put into evidence. The variance of estimation is then derived from the properties of the statistical estimator. The formulation is validated numerically, by comparison with the ECCDF variances derived from the synthetic data. The variance of the fluctuations around the CCDF is then shown to be dependent on the average of the correlation function of the time series, on the probability level and on the measurement duration. This variance of estimation is needed as a prerequisite in conjunction with the knowledge of the climatic variability to characterize the yearly fluctuations of propagation statistics computed from experimental time series. The extensions from simulations to experiments as well as the application to system planning are detailed in part II. Copyright © 2013 John Wiley & Sons, Ltd.

Received 11 January 2013; Revised 22 May 2013; Accepted 18 September 2013

KEY WORDS: rainfall rate statistics; rain attenuation statistics; empirical complementary cumulative distribution function; statistical estimator; inter-annual variability; confidence intervals

1. INTRODUCTION

Different Earth-space propagation experiments above 10 GHz such as SIRIO [1], ETS-II [2], COMSTAR [3], OLYMPUS [4], ITALSAT [5] and ACTS [6] have been carried out in Europe and in the USA allowing a significant amount of rain attenuation time series to be produced. Once processed and analysed to generate statistics, significant fluctuations have been found on the empirical complementary cumulative distribution functions (ECCDFs) of rain attenuation computed on a yearly basis [7, 8]. This variability of the ECCDFs is particularly noticeable when plotting the results in a log-scale for the probabilities, which is a convenient representation to get the propagation margin, which needs to be implemented to reach the required level of system availability.

Besides, a similar variability has also been reported on rainfall rate statistics generated from rainfall rate measurements [9–12]. For both rainfall rate and rain attenuation ECCDFs, the magnitude of the observed variability can be extremely high. For example, the ECCDFs computed on a yearly basis from the beacon measurements in Spino d’Adda for each of the 7 years of the ITALSAT experiment at
18.7 GHz [13] exhibit variations in the attenuation exceeded for $10^{-2}\%$ of the time ranging from 6 to 18 dB. The rainfall rate exceeded $10^{-2}\%$ of the time for the same years ranges from 25 to almost 100 mm/h. Such fluctuations have also been reported in [8] and in many other papers. A study of this variability has already been conducted in [7] and [8] in which an empirical model of variability, inferred from a limited collection of yearly experimental ECCDFs, has been proposed in order to predict the risk associated with propagation margins. So far, the reasons of this variability have been mainly related to climatic variability [11, 14] and also to the up-time of an experiment [15].

If these factors clearly contribute to the observed variability, they may not constitute the main source of fluctuations especially at low probability levels. Indeed, it is shown in the present paper that the statistical estimator used to compute the ECCDFs has a significant variance due to the finite duration of the measurements, to the low number of samples considered for low probability levels and to the correlation between successive samples. This correlation between the successive samples can be intuited as representative of the variations induced by the weather at short term, that is, for time lags of some hours or days. A proper quantification of this variability of empirical statistics is of great interest as it may enable the confidence level (expected deviations around the ‘long-term’ distribution) associated with experimental statistics covering a given duration to be estimated. As an example, for operational purposes, when designing a link for a given availability, the knowledge of this variability would allow the estimation of the magnitude of the deviations of the actual availability from the planned one as already mentioned in [8] or [9]. This issue could be critical (especially at Ka and Q/V bands) for satellite operators if the availability of a link is monitored and guaranteed from a contractual point of view. These issues will be discussed further in part II [16].

The paper is organized as follows. First, in Section 2, the variability of yearly ECCDFs around the rain attenuation complementary cumulative distribution function (CCDF) is highlighted using a rain attenuation time series synthesizer. The choice to work with synthetic data is motivated by the objective to quantitatively assess the ECCDF variability due to the statistical estimator only, disregarding other sources of fluctuations such as longer-term climatic variability (i.e. inter-annual, for instance) or uncertainties, which are usually present in experimental data. Then, in Section 3, the ECCDF variance of estimation is derived considering uncorrelated samples and correlated samples of the time series. The analytical formulations are then validated numerically, by comparison with the ECCDF variances derived from the synthetic data. Section 4 is a discussion that also gives the conclusions and perspectives of applications of the present work: weaknesses and assumptions of the approach are discussed as well as the way forward to include the inter-annual climatic fluctuations.

2. NUMERICAL SIMULATION OF THE IMPACT OF THE OBSERVATION DURATION ON THE VARIABILITY OF RAIN ATTENUATION STATISTICS

In order to stress the statistical variability of ECCDFs generated from rain attenuation data, ECCDFs computed from synthesized rain attenuation time series have been used. As mentioned in Section 1, this choice is motivated by different factors. First, the synthetic data do not include any climatic variability as the correlation between the attenuation samples is 0 after some hours. Second, there is no missing data due to experimental problems such as loss of lock, electric failures, limited range of the receiver and so on. Third, it enables to get very long datasets whereas the maximum duration of consecutive beacon measurements is ‘only’ of 7 years for the ITALSAT experiment in Spino d’Adda (which is the longest experiment carried out so far at Ka and Q/V bands, [13, 16]).

Therefore, even if the synthetic rain attenuation time series do not have exactly the same statistical properties as the experimental ones, their use to put into evidence the high variability associated with empirical statistics is expected to give an appropriate order of magnitude as their parameterizations are derived from measured data. In the present paper, the channel model used to generate the synthetic rain attenuation time series is described in [17]. As for the model of ITU-R Rec. P.1853 ([18, 19]), the approach relies on the conversion of a correlated stationary Gaussian process $G(t)$ into a rain attenuation process $A(t)$. The correlation function $\rho(t)$ of the Gaussian process $G(t)$ is set so that, after transformation, the correlation of $A(t)$ reproduces the correlation derived from experimental rain attenuation time series. Considering various rain attenuation experimental databases, [19] proposed the average formulation for the correlation function of the Gaussian process $\rho(t)$:

Copyright © 2013 John Wiley & Sons, Ltd.
\[ c_G(\tau) = \exp(-\beta \tau), \]  

where \( \tau \) is the time lag and \( \beta = 2 \times 10^{-4} \text{s}^{-1} \) (\( \beta^{-1} = 5000 \text{s} \)) defines the correlation time. This value is used in the following only for illustrative purposes, and different values for \( \beta \) would lead to slightly different results but with the same order of magnitude.

The attenuation process \( A(t) \) is finally obtained from a monotonic transformation of the standard centred Gaussian process \( G(t) \):

\[ A(t) = f(G(t)), \]

where \( f \) depends on the CCDF of rain attenuation. The random process \( A(t) \) obtained that way is ergodic because of the ergodicity of \( G(t) \). Consequently, for an infinite duration, the fraction of time spent by the rain attenuation process \( A(t) \) over a given threshold \( A^* \) (or ECCDF) computed from one realization of \( A(t) \) converges towards the probability \( P(A > A^*) \) that the rain attenuation \( A \) is over the threshold \( A^* \). If the convergence is ensured for an infinitely long period, deviations are expected for finite durations.

To illustrate this point, 500 years of rain attenuation time series sampled every second have been generated from the channel model [17], considering an arbitrary rain attenuation CCDF such that \( P(A > 0) = 5\% \) (a typical value for European climates). The dispersion of the ECCDFs computed for durations arbitrarily restricted to 1, 2, 5 and 10 years is illustrated on Figure 1. The grey lines represent the ECCDFs that can be derived from the 500-year time series. The black line is the CCDF.

![Variability of ECCDFs of Attenuation](image1.png)

**Figure 1.** Rain attenuation empirical complementary cumulative distribution functions (ECCDFs) derived from 500 years of simulation for durations restricted to (a) 1 year (500 ECCDF curves), (b) 2 years (250 ECCDF curves), (c) 5 years (100 ECCDF curves) and (d) 10 years (50 ECCDF curves). The sampling time is 1 s.
given as an input and plotted for reference. It is undistinguishable from the ECCDFs derived from the 500-year time series. Moreover, for successive attenuation thresholds $A^*$, the standard deviation $\sigma$ of the associated probability $p = P(A > A^*)$ is computed from the ECCDFs derived for the various durations. It is represented in Figure 1 by the $\pm \sigma$ curves around the CCDF.

Figure 1 clearly illustrates the variability of the ECCDFs around the CCDF. Two relatively obvious trends can be observed from Figure 1. First, the magnitude of the fluctuations increases for high attenuation thresholds (or low time percentages $p$) so that $\sigma = \sigma(p)$. Second, the magnitude of the fluctuations for a given attenuation threshold (or time percentage) decreases when the observation duration increases.

Considering short observation durations and high attenuation thresholds, the dispersion around the CCDF is extremely high. This has two main practical consequences. First, the modelling of the CCDF from empirical ECCDFs based on short observation durations and with a significant weight on the low probability thresholds is unlikely to be accurate because the information considered for the modelling process exhibits significant deviations from the CCDF. Second, if a link is designed to insure a given service availability for a finite duration from the CCDF, an additional safety margin is required to reduce the risk inherent to the statistical variability. A quantification of this variability is thus required to define advisedly the additional margin. To this aim, different statistical properties of ECCDFs are established and compared with the numerical simulation results in Section 3.

3. CHARACTERIZATION OF THE VARIABILITY OF EMPIRICAL COMPLEMENTARY CUMULATIVE DISTRIBUTION FUNCTIONS

3.1. Empirical complementary cumulative distribution function

Let $a_{i=1..N}$ be the rain attenuation (or rainfall rate) values collected during a propagation experiment. Assuming that the samples $a_{i=1..N}$ are realizations of the identically distributed random variables $A_{i=1..N}$, the ECCDF $\omega_N(A^*)$ of the measured rain attenuation (or rainfall rate) evaluated at the level $A^*$ is by definition given by ([20]):

$$\omega_N(A^*) = \frac{1}{N} \sum_{i=1}^{N} I(A_i > A^*),$$

(3)

where $I$ is the indicator function equal to 1 when $A_i > A^*$ and equals to 0, otherwise. Obviously, (3) is the number of collected samples that are over the threshold $A^*$ divided by the overall number of samples $N$. Assuming that the rain attenuation (or the rainfall rate) random variables $A_i$ are identically distributed random variables, the Glivenko–Cantelli theorem gives a convergence almost surely of the ECCDF $\omega_N(A^*)$ towards the CCDF $P(A > A^*)$ [20]. Therefore, the ECCDF is a consistent and unbiased estimator of the CCDF [21].

Considering that a sufficient number of samples are collected during 1 year of experiment; for instance, $N$ is often thought to be sufficiently large so that the deviations of the ECCDF around the CCDF are assumed to be low. Both assumptions are often mistaken. Indeed, the ECCDF $\omega_N(A^*)$ is a random variable—as it is a sum of random variables in compliance with (3)—and has consequently a variance. It is shown in Section 3.2 (uncorrelated samples) and especially in Section 3.3 (correlated samples) that this ECCDF variance of estimation is the cause of the fluctuations reported on Figure 1.

3.2. Empirical complementary cumulative distribution function variance of estimation for uncorrelated samples

Each term $I(A_i > A^*)$ introduced in (3) defines a Bernoulli random variable $U_i$ equal to 1, if $A_i > A^*$ and equal to 0, otherwise. In such conditions,

$$\begin{align*}
P(U_i = 0) &= 1 - p \\
P(U_i = 1) &= p
\end{align*}$$

(4)

where $p = P(A > A^*)$. Denoting by $E[\cdot]$ the expectancy operator, it follows that $E[U_i] = p$ and $Var[U_i] = p(1 - p)$. In compliance with (3), the ECCDF for the threshold $A^*$ now reduces to $\omega_N(A^*) = \frac{1}{N} \sum_{i=1}^{N} U_i = U_N$.
and is the empirical average $\mathcal{U}_N$ of the random variables $U_i$. Note that $\mathcal{U}_N$ is also a random variable and it is straightforward to show that the expectation of the ECCDF is $E[\mathcal{U}_N] = p$, which confirms that this estimator is unbiased. The variance of estimation $\sigma_E^2 = Var[\mathcal{U}_N]$ of the ECCDF is then given by

$$\sigma_E^2 = \frac{1}{N^2} E\left[ \sum_{1 \leq i, j \leq N} U_i U_j \right] - p^2$$

(5)

As the samples are assumed to be independent, it follows that $E[U_i U_j] = E[U_i] E[U_j] = p^2$ for $i \neq j$. Moreover, recalling that $Var[U_i] = E[U_i^2] - E[U_i]^2$, it follows that $E[U_i^2] = p(1 - p) + p^2 = p$ and the variance of estimation $\sigma_E^2 = Var[\mathcal{U}_N]$ driven by (5) finally reduces to

$$\sigma_E^2 = \frac{1}{N^2} \left( (N^2 - N)p^2 + Np \right) - p^2$$

(6)

In compliance with (6), the variance of estimation $\sigma_E^2 = Var[\mathcal{U}_N]$ of the ECCDF depends on the probability $p = P(A > A^*)$ to exceed $A^*$ and on the number of samples $N$ so that $\sigma_E^2 = \sigma_E^2(p, N)$. In addition, the central limit theorem states that the distribution of $\mathcal{U}_N$ converges towards a normal distribution with average $p$ and variance $\sigma_E^2$, if $N$ is sufficiently large. To assess the impact of the variance of estimation $\sigma_E^2$, it is convenient to refer to the 68% confidence interval that corresponds to plus or minus one standard deviation around the probability for a normally distributed quantity. Considering (6), the confidence interval at 68% is given by $[p - \sigma_E, p + \sigma_E]$. The half width of this 68% confidence interval normalized to $p$ (i.e. $\sigma_E(p, N)/p$) is illustrated on Figure 2 as a function of the probability level $p$, considering a fictitious experiment lasting 1 year with different sampling times varying from 1 to 3600 s (resulting thus in various number of samples $N$).

In compliance with Figure 2, considering, for instance, an attenuation level exceeded with a probability $p = 10^{-2}$%, $\sigma_E(p, N)/p$ (in %) ranges from 1.5% if $N \approx 3 \times 10^7$ samples are considered (1 s sampling time) to 45% for $N = 52,500$ samples (600 s sampling time).

This confidence interval is also reported on Figure 3 as a function of the sampling time (i.e. $N$) around the rain attenuation CCDF used in Section 2. According to Figure 3, there is a probability of 32% (i.e. 100–68%) that the ECCDF computed from 1 year of measurements with a given sampling time lies outside the associated 68% confidence interval. On the one hand, considering that the sampling time of attenuation measurements is usually around 1 s, the estimator variability could be

![Figure 2](image-url)
seen as a minor issue. On the other hand, considering that the sampling time of rainfall rate measurements is usually about some minutes, the variability is larger but not dramatic. Nevertheless, as shown in Figure 3, the variability increases rapidly with the decrease of the probability level \( p \), but the low magnitude of the fluctuations for samples every 1 s does not compare satisfactorily with the fluctuations reported for 1 year in Figure 1. Now, it has to be kept in mind that Figure 1 derives from the simulation of a correlated rain attenuation process, whereas the results here have been established assuming independence between the samples. This assumption is not fulfilled in most of the experimental configurations because successive samples of rainfall rate or rain attenuation are not uncorrelated. The effect of this correlation on the ECCDF variability is assessed in Section 3.3 and is shown to explain the fluctuation discrepancies between Figures 3 and 1.

### 3.3. Empirical complementary cumulative distribution function variance of estimation for correlated samples

#### 3.3.1. Analytical derivations

To account for the correlation effects on the ECCDF variability, it is now considered that the time-series collected during an experiment are one realization of a correlated stationary discrete random process. In such conditions, the \( N \) samples collected during an experiment are now realizations of the correlated random variables now denoted \( A(\Delta t)_i = 1 \) if \( A(i\Delta t) > A^* \), \( U(i\Delta t) = 0 \) otherwise is defined in compliance with Section 3.2. The random process \( U(t) \), because of the stationarity of the process \( A(t) \), is stationary so that its correlation function \( c_U \) depends only on the time lag \( |i - j|\Delta t \). The estimator of the CCDF can then be rewritten as \( \omega_{\text{CCDF}}(A^*) = \frac{1}{N} \sum_{i=1}^{N} U(i\Delta t) = U_N \).

The variance of estimation \( \sigma_E^2 = \text{Var}[\overline{U}_N] \) of the ECCDF for a given threshold \( A^* \), still given by the general form (5), now becomes, for correlated samples,

\[
\sigma^2_E = \text{Var}[\overline{U}_N] = E[\overline{U}_N^2] - E[\overline{U}_N]^2
\]

\[
= E\left[ \frac{1}{N} \sum_{i=1}^{N} U(i\Delta t) \right]^2 - p^2
\]

\[
= \frac{1}{N^2} \sum_{1 \leq i,j \leq N} E[U(i\Delta t)U(i\Delta t)] - p^2
\]

The correlation function \( c_U \) of the stationary process \( U(t) \) is by definition
\[ c_U((i-j)\Delta t) = \frac{E[U(i\Delta t)U(j\Delta t)] - E[U]^2}{\text{Var}[U]}. \] (8a)

Recalling that \( E[U] = p \) and \( \text{Var}[U] = p(1-p) \), (8a) becomes

\[ c_U((i-j)\Delta t) = \frac{E[U(i\Delta t)U(j\Delta t)] - p^2}{p(1-p)} = c_U((i-j)\Delta t, p), \] (8b)

which underlines the implicit dependency of \( c_U \) with respect to the probability \( p = P(A > A^\circ) \). It follows that (7) can be turned into

\[
\sigma^2_{E} = \text{Var}[\bar{U}_N] = \frac{1}{N^2} \sum_{i \neq j \in \mathbb{N}} (c_U((i-j)\Delta t, p)\text{Var}[U] + p^2) - p^2 \\
= \frac{\text{Var}[U]}{N^2} \left[ Nc_U(0) + 2 \sum_{i=1}^{N-1} (N-i)c_U(j\Delta t, p) \right] \\
= \frac{\text{Var}[U]}{N^2} \left[ \sum_{i=-N+1}^{N-1} \left( 1 - \frac{|i|}{N} \right)c_U(j\Delta t, p) \right] \\
= \frac{p(1-p)}{N} \sum_{i=-N+1}^{N-1} \left( 1 - \frac{|i|}{N} \right)c_U(j\Delta t, p) \\ (9)
\]

According to (9), the variance of estimation \( \sigma^2_{E} = \text{Var}[\bar{U}_N] \) of the ECCDF still depends not only on the probability level \( p \) and on the number of samples \( N \) but also on the correlation function \( c_U \) of the random process \( U(t) \) so that now \( \sigma^2_{E} = \sigma^2_{E}(p, N, c_U) \). Importantly, (9) is consistent with (6) considering independent samples (i.e. \( c_U(0) = 1, c_U(\tau) = 0, \) otherwise).

In accordance with (9), the evaluation of the ECCDF variance of estimation for rain attenuation or rainfall rate processes as a function of the probability level \( p \) requires the definition of the correlation function \( c_U(\tau) = c_U(\tau, p) \) of the Bernoulli process \( U(t) \). This can be performed for the synthesizer used in Section 2. Indeed, the approach proposed in Rec. ITU-R P.1853 to simulate rain attenuation time series lies first in the generation of a correlated Gaussian process \( G(t) \) with correlation \( c_G(\tau) = \exp(-\beta\tau) \), where \( \beta = 2 \times 10^{-4} \text{s}^{-1} \). Underlying Gaussian process \( G(t) \) is then turned into a rain attenuation process \( A(t) \) using the monotonic function \( f \) so that \( f(G(t)) = A(t) \) (see [17, 18] or [19] for more details on the mathematical definition of the rain attenuation time series synthesizer). In such conditions, the Bernoulli process definition \( U(i\Delta t) = 1 \) if \( A(i\Delta t) > A^\circ \) and \( U(i\Delta t) = 0 \) otherwise is strictly equivalent to \( U(i\Delta t) = 1 \) if \( G(i\Delta t) > \alpha \) and \( G(i\Delta t) = 0 \) otherwise where \( \alpha = f^{-1}(A^\circ) \).

As \( A^\circ \) is defined so that \( E[U(t)] = p \), it follows that \( E[I(G(t) > \alpha)] = P(G(t) > \alpha) = p. \) Recalling that \( G(t) \) is a reduced centred Gaussian process, the threshold \( \alpha \) is obtained through \( \alpha = \sqrt{2\text{erfc}^{-1}(2p)} \), where \( \text{erfc}^{-1} \) is the inverse complementary error function. Therefore, whatever the probability level \( p \), the Bernoulli process \( U(t) \) can be related to the Gaussian process \( G(t) \) through the transformation \( \Phi \)

\[ U(t) = \Phi(G(t), p), \] (10)

with

\[
\Phi(x, p) = \begin{cases} 
0 & x < \sqrt{2}\text{erfc}^{-1}(2p) \\
1 & x \geq \sqrt{2}\text{erfc}^{-1}(2p) 
\end{cases} \\ (11)
\]

From (10) and (11), the covariance \( \text{Cov}[U](\tau) \) of the binary process \( U(t) \)—and then \( c_U(\tau) \) from (8a) and (8b)—can now be expressed as a function of the correlation \( c_G(\tau) \) of the underlying Gaussian process \( G(t) \) by [22, 23]

\[
\text{Cov}[U](\tau) = \sum_{k=1}^{\infty} \frac{\Phi_k^2(p)}{k!} c_G(\tau)^k, \\ (12)
\]

where \( \Phi_k(p) \) are coefficients related to the projections of the function \( \Phi(x, p) \) on the orthogonal basis of the Hermite polynomials. Further details on the methodology can be found in the Appendix. Using the expression of the covariance (12) allows deriving the correlation \( c_U(\tau, p) \) through (8b). It is illustrated
on Figure 4 where evaluations of \( c_U(\tau,p) \) for different probability levels \( p \) are given. Now that \( c_U(\tau,p) \) is known, it is a simple matter to compute the variance of estimation \( \sigma_E^2 \) in (9).

3.3.2. Effect of the sampling time and the observation duration on the empirical complementary cumulative distribution function variance of estimation. The ECCDF variance of estimation \( \sigma_E^2 \) has been computed according to (9) for different sampling times \( \Delta t \) considering 1 year of acquisition. First, the normalized width of the 68% confidence interval \( \sigma_E(p,N)/p \) introduced in Section 3.2 is reported on Figure 5 for different sampling times \( \Delta t = 1, 10, 60, 300, 600, 3600 \) s. The 68% confidence interval \( \pm \sigma_E(p,N) \) is reported on Figure 6 around the rain attenuation CCDF already used in Section 2 to generate the synthetic data.

Figures 5 and 6 respectively differ from Figures 2 and 3, where uncorrelated samples were considered. Particularly, for correlated samples, the ECCDF variance of estimation \( \sigma_E^2 \) is clearly higher for the lowest probability levels. \( \sigma_E^2 \) seems to be insensitive to the sampling time \( \Delta t \) whenever \( \Delta t \) is small and \( N\Delta t \) large with respect to the correlation time \( \beta^{-1} = 5000 \) s. This result has to be expected. Indeed, recalling that \( c_U(k\Delta t) \) tends towards 1 for \( k\Delta t \ll \beta^{-1} \) and towards 0 for \( k\Delta t \gg \beta^{-1} \) in compliance with Figure 4, it follows that

\[
\frac{1}{N} \sum_{k=-N+1}^{N-1} \left( 1 - \frac{|k|}{N} \right) c_U(k\Delta t) = \Sigma/N \quad \text{in (9) can be approximated by}
\]

![Figure 4. Correlation functions \( c_U(\tau) \) derived from (12), (8a) and (8b) for different probability levels \( p \). The underlying Gaussian process has a correlation function \( c_G(\tau) \) given by (1) with \( \beta = 2 \times 10^{-4} \) s\(^{-1} \).](image)

![Figure 5. Normalized half width of the 68% confidence interval (in per cent of the probability level) derived from (9) as a function of the probability level \( p \), considering 1 year of correlated samples with different sampling times varying from 1 to 3600 s.](image)
\[ \Sigma \approx \frac{1}{N} \sum_{k=-N+1}^{N-1} c_U(k\Delta t) \]. In such conditions, \( \Sigma/N \) in (9) reduces to the average (or the integral) of the correlation function \( c_U \) using a Riemann sum. Therefore, as long as the summation step driven by \( \Delta t \) is sufficiently small with regard to \( \beta^{-1} = 5000 \), the sum is well approximated, and the sampling time has a very low impact. By analogy with the uncorrelated case developed in Section 3.2, that is, comparing (6) and (9), \( \frac{1}{N} \sum_{k=-N+1}^{N-1} \left( 1 - \frac{|k|}{N} \right) c_U(k\Delta t) \) can be regarded as the inverse of the number of uncorrelated samples used to compute the statistics. Considering 1 year of samples in the correlated case, the equivalent number of independent samples ranges from \( \sim 10000 \) for 1% of a year to \( \sim 24000 \) for 10% of a year considering sampling times ranging from 1 to 600 s. The effect of the correlation can then be understood intuitively as a reduction of the amount of information provided by individual samples.

In compliance with (9), whenever \( \Delta t \) is fixed and negligible with respect to \( \beta^{-1} \), the only way to reduce the variance of estimation \( \sigma_E^2 \) is to increase the duration of the time series with respect to the correlation length of the process (i.e. to increase the number of samples \( N \)). This point is illustrated in Figures 7 and 8 where observation periods ranging from 1 to 100 years have been considered. It can then be observed that the variance of estimation decreases when the time series duration increases.

Figure 6. Confidence interval at 68% as a function of the probability level \( p \) for different sampling times \( \Delta t \) considering correlated samples and 1 year of data.

Figure 7. Normalized half width of the 68% confidence interval (in per cent of the probability level) derived from (9) as a function of the probability level \( p \), considering observation durations ranging from 1 to 100 years. The sampling time \( \Delta t \) is fixed to 60 s.
Assuming the integrability (or summability) of the correlation function \( c_U \), there exist a sufficient number of samples \( N \) so that

\[
\sum_{i=-N+1}^{N-1} \left( 1 - \frac{|i|}{N} \right) c_U(i \Delta t) \approx \lim_{N \to \infty} \sum_{i=-N+1}^{N-1} c_U(i \Delta t) = \Sigma. \tag{13}
\]

Considering a sufficient duration \( T = N \Delta t \) for the observations, the variance of the estimation \( \sigma_E^2 \) in (9) reduces to

\[
\sigma_E^2 = \frac{p(1-p)K}{N} = \frac{p(1-p)\Sigma \Delta t}{T}. \tag{14}
\]

From a practical point of view, (14) is valid whenever the observation duration \( T \gg \beta^{-1} \). As shown in Figure 8, this demonstrates that the ECCDF standard deviation \( \sigma_E \) at a given probability level \( p \) is inversely proportional to the square root of the observation duration. The existence of the limit \( \Sigma \) in (13) is intrinsically linked to the ergodicity of the process and hence to the consistency of the ECCDF estimator. In the case of the synthetic rain attenuation times series considered in the present paper, this limit exists because of the exponential decay of the correlation function \( c_U \) reported on Figure 4. In the case of natural rain attenuation or rainfall rate processes, the existence of the limit is far from being evident as fractals or other long range dependence processes have been proposed to model rainfall rate or rain attenuation processes [24]. Nevertheless, from experimental rain attenuation time series, it will be shown in part II [16] that an exponential decaying function is an acceptable representation for \( c_U \).

### 3.3.3. Numerical verification of the results.

In order to check the validity of the statistical framework laid in Section 3.3.1, the standard deviation \( \sigma_E \) of the ECCDF driven by (9) is compared with the standard deviation of the ECCDFs derived from the numerical simulations conducted in Section 2. The impact of the measurement duration is verified on Figure 9, whereas the impact of the sampling time \( \Delta t \) is checked on Figure 10.

Figures 9 and 10 confirm the validity of the formalism laid in Section 3.3.1 to evaluate the variance of estimation of the ECCDFs. The slight departures that can be observed in Figure 10 between the numerical and theoretical \( \pm \sigma_E \) curves are related to the low number of synthetic samples obtained for the lower probabilities and the higher sampling time.

### 4. DISCUSSION AND PROSPECTS

It has been demonstrated that the statistical estimator used to compute rain attenuation or rainfall rate ECCDFs from time series has a variance \( \sigma_E^2 \), which derivation has been conducted. The analytical
formulation shows that \( \sigma^2_E \) depends on the probability level \( p \), on the short term correlation function \( c_U \) related to the rainfall rate or rain attenuation process and on the duration (number of samples \( N \)) of the time series from which the ECCDFs are computed.

Figure 9. Standard deviations \( \pm \sigma_E \) of the empirical complementary cumulative distribution function computed from (9) compared with the standard deviation derived from the numerical simulations conducted in Section 2. The observation durations range from 1 to 10 years, the sampling time \( \Delta t \) is 1 s.

Figure 10. Effect of the sampling period on the 68% confidence interval \( \pm \sigma_E \) associated to an empirical complementary cumulative distribution function computed from (9) and from the numerical simulations conducted in Section 2.
To validate the mathematical framework and to quantify the statistical variability but without loss of generality, synthetic rain attenuation time series have been considered. It has then been shown that the variance of estimation $\sigma^2_E$ of the ECCDFs is far from negligible if the statistics of interest are derived from yearly time series (as it is commonly the case) and can be especially high for low probability levels usually considered to design fixed Earth-space links (commonly between 99.9% and 99.99% of the time).

In order to obtain results relevant for link design, the variance of estimation $\sigma^2_E$ given by (9) must now be parameterized and validated from experimental measurements. Particularly, (9) requires the definition of the correlation $c_\omega(t,p)$ of the Bernoulli process $U(t)$ as a function of the probability $p$. Considering synthetic rain attenuation time series, which first and second order statistical properties are known, has made this derivation possible. Now, without questioning the validity of rain attenuation time series synthesizers such as [17, 18] and [19], much more significant results are expected if $c_\omega(t,p)$ is derived directly from experimental data. This is precisely what is addressed in part II [16].

In other respects, ECCDFs are usually computed from yearly experimental time series of rain attenuation or rainfall rate. If the variance of estimation is derived directly from experimental data. This is precisely what is addressed in part II [16].

As an illustration, assume that an error-free model of the rain attenuation CCDF $P_y(\omega)$ contains the two causes of variability of the yearly ECCDFs, namely the estimation noise $\sigma^2_E$ and the climate-induced fluctuations $\sigma^2_C$ of the probability distribution, which derivation is addressed in part II [16].

The knowledge of the overall variability $\sigma^2(\omega)$ is of prime importance for Earth-space link design. As an illustration, assume that an error-free model of the rain attenuation CCDF $P_y(\omega)$ is at disposal (or with a reasonable estimation of its accuracy) and consider a finite duration. The probability that the measured availability is lower than the one for which the system was designed is of 50%. An additional margin is thus required to reduce this risk of being under the specifications. Its definition requires the knowledge of the variability $\sigma^2(\omega)$ around the probabilities $p$, for the considered duration. This subject may become highly critical as for Ka and Q/V bands broadband satellites, the operators can monitor and compute statistics on the link state as the carrier to noise plus interference ratio of...
the forward link is returned to the gateways for the choice of the most appropriate physical layer. Therefore, the emergence of contractual terms between satellite operators and their customers, guaranteeing a measured availability, is technically possible. These aspects, as well as the parameterization and validation of \( \sigma^2_E(p) \) and \( \sigma^2_C(p) \) from real data, are detailed in part II [16].

APPENDIX

This appendix aims at providing the derivation of the binary process correlation function \( c_U(\tau) \) from the correlation \( c_G(\tau) \) of the underlying Gaussian process. Consider two random processes \( X(t) \) and \( G(t) \) such that \( X(t) = \Phi(G(t)) \). Assuming that \( G(t) \) is a stationary Gaussian process with correlation function \( c_G(\tau) \), the covariance \( \text{Cov}[X](\tau) \) of the process \( X(t) \) is related to \( c_G \) by ([20])

\[
\text{Cov}[X](\tau) = \sum_{k=1}^{\infty} \frac{\Phi_k^2}{k!} [c_G(\tau)]^k \tag{A1}
\]

with \( \Phi_k = \int_{-\infty}^{\infty} H_k(x) \phi(x) g(x) dx \) and where \( H_k(x) \) is the Hermite polynomial of degree \( k \) and \( g(x) \) is the unit Gaussian measure. The Hermite polynomials are defined by ([22])

\[
H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2}. \tag{A2}
\]

The following recurrence relationship holds and can be used to compute the polynomials of higher order:

\[
H_{n+1}(x) = xH_n(x) - nH_{n-1}(x) \tag{A3}
\]

In the present paper, the Gaussian process \( G(t) \) is transformed into a binary process \( U(t) \) according to

\[
\begin{cases}
\Phi(G) = 0 & G \leq \alpha \\
\Phi(G) = 1 & G > \alpha
\end{cases} \tag{A4}
\]

where \( \alpha = \sqrt{2} \text{erfc}^{-1}(2p) \). Thus, the coefficients \( \Phi_k \) can be expressed as

\[
\Phi_0 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{1}{2} \text{erfc} \left( \frac{\alpha}{\sqrt{2}} \right) = p \tag{A5}
\]

\[
\Phi_1 = \int_{-\infty}^{\infty} \frac{d}{dx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \frac{e^{-\alpha^2/2}}{\sqrt{2\pi}} \tag{A6}
\]

\[
\Phi_k = \int_{-\infty}^{\infty} (-1)^k \frac{1}{\sqrt{2\pi}} \frac{d^k}{dx^k} e^{-x^2/2} dx
= \frac{1}{\sqrt{2\pi}} \left[ (-1)^{k-1} \frac{d^{k-1}}{dx^{k-1}} e^{-x^2/2} \right] (\alpha)
= \frac{e^{-\alpha^2/2}}{\sqrt{2\pi}} H_{k-1}(\alpha) \tag{A7}
\]

Considering the recurrence relationship existing between the \( H_k \), it comes

\[
\Phi_{k+1} = \alpha \Phi_k - (k-1) \Phi_{k-1} \tag{A8}
\]

In order to perform a numerical computation, a large number of terms of the series (\( \approx 5000 \)) have to be computed, and the recurrence should be made on \( \psi_k = \frac{\Phi_k}{\sqrt{k!}} \) to avoid issues linked to the overflow of the maximum single precision floating point value. It gives
\[ \psi_{k+1} = \frac{a}{\sqrt{k+1}} \psi_k - \frac{k-1}{k(k+1)} \psi_{k-1} \]

It enables to compute the series (A1) extremely efficiently with a good accuracy (the maximum of the convergence error is bounded by the difference between 1 and the estimated correlation in 0 giving thus a mean to check the convergence). Considering the correlation function (1), with \( \beta = 2 \times 10^{-4} \, s^{-1} \), the correlation functions of the binary processes for various probability levels \( p \) can be computed as illustrated on Figure 4.

Acknowledgment

This activity has been partly carried out in the framework of the European action COST IC-0802.

References


Authors’ Biographies

Nicolas Jeannin works as a research engineer in the radio communication and propagation group of ONERA since 2009. He obtained his PhD in the field of tropospheric channel modelling from the University of Toulouse in 2008 and his Engineering Degree and MSc from ISAE in the field of remote sensing in 2005. His current research interest concerns the impact and the modelling of the troposphere on satellite telecommunication systems and remote sensing instruments.
Xavier Boulanger received his Diploma of Engineering in Electronics and Digital Communications from École Nationale de l’Aviation Civile (ENAC), Toulouse, France, in 2010. He received his PhD degree in propagation and signal processing in 2013 from Institut Supérieur de l’Aéronautique et de l’Espace (ISAE), Toulouse, France. In 2013, he joined the French Aerospace Lab (ONERA), where he is currently working on propagation channel modelling for Earth-space communication systems. He was involved in various working groups dealing with Earth-space propagation modelling such as COST IC0802, and ITU-R study group 3.

Laurent Castanet works as a research engineer in the radio communication field at ONERA Toulouse. His main research interests are Earth-space propagation (more particularly the dynamics and spatial variability of the propagation channel) and fade mitigation techniques (design and performance simulation). He is the head of the Radio-Communication and Propagation research unit of the Electromagnetics and Radar department of ONERA. He has been involved in propagation and radio communication studies for satellite systems at Ku, Ka and Q/V-bands. He is involved more particularly in studies related to the dynamics of the propagation channel, to the development and validation of propagation models in temperate and tropical areas as well as to fade mitigation techniques for Ka and Q/V band systems for CNES, ESA, European Industry and French MoD. He was a French expert in COST 255, COST 280 European projects, in which he worked on the development of advanced models of the space-time variability of the propagation channel and on the design and optimization of adaptive fade mitigation techniques. He represented ONERA in the European Network of Excellence ‘SatNEx’, in which he chaired the ‘Channel modelling and propagation impairment simulation’ Joint Activity. He is currently involved in the new European action COST-IC0802 in which he chairs the second working group devoted to ‘Channel modelling from C to W band’. He is also the head of the French Delegation to ITU-R Study Group 3, which deals with radiowave propagation. He teaches radiowave propagation in French Engineering Schools and University (ISAE, Telecom Bretagne, University Paul Sabatier) as well as link analysis at SAE Training institution.

Frédéric Lacoste received his MSc degree from École Nationale Supérieur d’Ingénieurs en Construction Aéronautique et Spatiale (ENSICA), Toulouse, France, in 2002 and his PhD degree in propagation and signal processing in 2005 from École Nationale Supérieure de l’Aéronautique et de l’Espace (SUPAERO), Toulouse, France. In 2005, he joined the French Space Agency (CNES), where he is currently working on propagation channel modelling for fixed and mobile Earth-space communication systems. He was involved in various working groups dealing with Earth-space propagation modelling such as SatNEx I and II, COST 280 and, currently, COST IC0802, and ITU-R study group 3.