

Probabilistic Transient Stability Assessment and On-Line Bayes Estimation

Elio Chiodo and Davide Lauria

Abstract It is a well-known fact that the increase in energy demand and the advent of the deregulated market mean that system stability limits must be considered in modern power systems reliability analysis. In this chapter, a general analytical method for the probabilistic evaluation of power system transient stability is discussed, and some of the basic contributes available in the relevant literature and previous results of the authors are reviewed. The first part of the chapter is devoted to a review of the basic methods for defining transient stability probability in terms of appropriate random variables (RVs) (e.g. system load, fault clearing time and critical clearing time) and analytical or numerical calculation. It also shows that ignoring uncertainty in the above parameters may lead to a serious underestimation of instability probability (IP). A Bayesian statistical inference approach is then proposed for probabilistic transient stability assessment; in particular, both point and interval estimation of the transient IP of a given system is discussed. The need for estimation is based on the observation that the parameters affecting transient stability probability (e.g. mean value and variances of the above RVs) are not generally known but have to be estimated. Resorting to “dynamic” Bayes estimation is based upon the availability of well-established system models for the description of load evolution in time. In the second part, the new aspect of on-line statistical estimation of transient IP is investigated in order to predict transient stability based on a typical dynamic linear model for the stochastic evolution of the

The singular and plural of names are always spelled the same; boldface characters are used for vectors; random variables (RVs) are denoted by uppercase letters.

E. Chiodo (✉) and D. Lauria
Electrical Engineering Department, University of Naples Federico II, Via Claudio 21,
I-80125 Naples, Italy
e-mail: elio.chiodo@unina.it

D. Lauria
e-mail: davide.lauria@unina.it

system load. Then, a new Bayesian approach is proposed in order to perform this estimation: such an approach seems to be very appropriate for on-line dynamic security assessment, which is illustrated in the last part of this article, based on recursive Bayes estimation or Kalman filtering. Reported numerical application confirms that the proposed estimation technique constitutes a very fast and efficient method for “tracking” the transient stability versus time. In particular, the high relative efficiency of this method compared with traditional maximum likelihood estimation is confirmed by means of a large series of numerical simulations performed assuming typical system parameter values. The above results could be very important in a modern liberalized market in which fast and large variations are expected to have a significant effect on transient stability probability. Finally, some results on the robustness of the estimation procedure are also briefly discussed in order to demonstrate that the methodology efficiency holds irrespective of the basic probabilistic assumptions made for the system parameter distributions.

List of main symbols and acronyms

BCI	Bayesian confidence interval
cdf	Cumulative distribution function
CSGDF	Complementary standard Gaussian distribution function
CCT	Critical clearing time (T_{cr} or T_x)
CV	Coefficient of variation
D	Set of observed data used for inference
DLM	Dynamic linear model
$E[R]$	Expectation (or “mean value”) of the RV R
EV	Extreme value distribution
FCT	Fault clearing time
$F(x)$	Generic cdf
$f(x)$	Generic pdf
$g(\omega)$, $g(\omega D)$	Prior and posterior pdf of a generic parameter ω
$G(r, \phi)$	Gamma distribution with parameters (r, ϕ)
IID	s-Independent and identically distributed (random variables)
IP	Instability probability
LCCT	Logarithm of the CCT
LF, $L(D \beta)$	Likelihood function, conditional to given parameter β
L , $L(t)$	Load (at time t)
ML	Maximum likelihood
MSE	Mean square error
$LN(\alpha, \beta)$	Log-Normal distribution with parameters α and β
$N(\mu, \sigma)$	Normal (Gaussian) distribution with mean μ and SD σ
pdf	Probability density function
RV	Random variable
s	Standard deviation of measurement errors in the DLM of the LCCT
S	Standard deviation of measurement errors in the DLM of the load

SD, σ	Standard deviation
s -independent	Statistically independent
SD[Y]	Standard deviation of the RV Y
SM	Stability margin (i.e. the quantity u defined below), for a given fault
σ^2	Denotes a variance
T_{cr} or T_x	Critical clearing time
T_{cl} or T_y	Fault clearing time
u	$(\alpha_x - \alpha_y)/(\beta_x^2 + \beta_y^2)^{1/2}$
v_x, v_y	CV values of the CCT and FCT, respectively
VST	Very short time
Var[R], $V(R)$	Variance of the RV R
w	Standard deviation of system equation error in the DLM of the LCCT
W	Standard deviation of system equation error in the DLM of the load
Λ	Peak value of the load $L(t)$, over a given time interval
WGN	White Gaussian noise
X	Logarithm of the CCT
Y	Logarithm of the FCT
α_x	$E[X]$
α_y	$E[Y]$
β_x^2	Var[X]
β_y^2	Var[Y]
ζ°	Bayes estimate of a generic parameter ζ
ζ^*	ML estimate of a generic parameter ζ
μ	Denotes a mean value (expectation)
$\hat{\mu}_k$	Bayes estimate of a “dynamic” parameter μ at time k
$\Gamma(\cdot)$	Euler–Gamma function
μ_r	Expectation of the generic RV R
$\Phi(z)$	Standard normal cdf
$\Psi(z)$	$1 - \Phi(z)$ (Complementary standard Gaussian distribution function)
$\varphi(z)$	Standard normal pdf
$R \sim N(\alpha, \beta)$	The RV R has a Gaussian distribution $N(\alpha, \beta)$ (and similarly for the LN model, etc.)

1 Introduction

Stability assessment has long been recognized as a fundamental requirement in power system planning, design, operation and control. Transient stability can be

defined as a property of an assigned power system to remain in a certain equilibrium point under normal conditions and reach a satisfactory equilibrium point after large disturbances such as faults, loss of generation, line switching, etc. [1]. Transient stability, therefore, constitutes a key aspect of modern power system reliability, and this fact is increasingly recognized in the modern power systems literature [2]. Indeed, some methods based on reliability theory are used in this chapter to perform an efficient assessment of power system stability. The traditional approach to transient stability analysis is deterministic, being based on the “worst case” approach. More specifically, transient stability quantitative assessment is generally performed on a three-phase fault on specific system buses as well as considering the load demand attaining its peak value over a prefixed time interval.

The application of probabilistic techniques for transient stability analysis was introduced in a series of articles by Billinton and Kuruganty [3–7], motivated by the random nature of:

- the system steady-state operating conditions;
- the time of fault occurrence;
- the fault type and location;
- the fault clearing phenomenon.

In fact, the steady-state operating conditions that heavily affect stability strongly depend on the load, which is a random process due to its intrinsic nature. This is especially evident in planning studies where the load level is the major source of uncertainty.

The time to clear the fault (fault clearing time, FCT), a crucial parameter in stability investigations, is also not known in advance, and so it should also be regarded as an RV. The probabilistic approach has also been explored in other significant articles such as [8], based on Monte Carlo simulations and [9], based on the “conditional probability” approach. An exhaustive account of the topic and the relevant bibliography can be found in the book by Anders [10, Chap. 12] which clearly states that: “stability analysis is basically a probabilistic rather than a deterministic problem”.

The analytical computation of the probability distributions of the intermediate RVs is one of the most challenging aspects due to the complexity of the mathematical models, as also pointed out by Anders [10, p. 577].

In a few articles [11–15], some theoretical results from probability theory and statistics have been utilized in order to develop an analytical approach to the transient stability evaluation of electrical power systems by performing critical considerations on the basic probability distributions. Instability probability (IP) over a certain period of time, with regard to a given fault, is defined and calculated by means of both critical clearing time (CCT) distribution and FCT distribution. This analytical approach, overcoming the drawbacks of Monte Carlo simulations, is very useful in actual operation since it permits straightforward sensitivity analysis of IP with regard to system parameters thus highlighting those which mostly influence the system stability characteristics and providing a quantitative

tool for performing proper preventive control actions. For similar reasons, the proposed analytical probabilistic approach is also a powerful tool with regard to the practical aspect of the estimation of the basic parameters relevant to the transient stability assessment like IP or other measures of “stability margin” (SM). This topic—generally neglected, or dealt with in approximate “sensitivity analyses” in the literature—was faced in [16], where an effort was made to tailor a simple, analytical, transient stability probability estimator which allows the required characteristics of both efficiency and robustness to be obtained, in the framework of classical estimation.

In this article, a new Bayesian approach is proposed in order to provide this estimation: such an approach appears to be the most suitable one for on-line transient stability assessment. Numerical application performed confirms that the estimation technique is able to adequately “track” the transient stability in time, being far more efficient than the classical maximum likelihood (ML) estimation of the IP. This could be an interesting property in a modern liberalized market in which fast and large variations are expected to have a significant effect on transient stability probability.

In the final part of the chapter, some results on the robustness of the estimation procedure are also briefly discussed in order to illustrate that the methodology efficiency holds irrespective of the basic probabilistic assumptions effected with regard to “a priori” distributions of the various system parameters.

In the four Appendices to the chapter:

1. a mathematical study of the IP versus the system parameters is illustrated in order to establish a proper “sensitivity analysis” of system stability which can be useful in the design stage.
2. some basic properties of Bayesian estimation, relevant for the problem under study, are briefly mentioned and some properties already derived by the authors in previous articles for the interval estimation of the IP are also included.

2 Probabilistic Modelling for Transient Stability Analyses

2.1 Definition and Evaluation of Transient IP

Many state variables of an electrical power system possess an intrinsically stochastic nature and, consequently, a probabilistic description of transient stability aspects is able to infer interesting deductions also in terms of control actions for improving system robustness. For instance, the steady-state conditions, fault conditions and circuit breaker clearing times are not precisely known or predictable. The various involved uncertainties should be properly taken into account using suitable probabilistic models. In a probabilistic frame, both the network configuration and faults are described as random quantities. According to this approach, all faults potentially causing instability and all the possible network

states at the instant of fault (e.g. in terms of the requested loads at load buses) have to be considered, together with their probability of occurrence. Once the cost of any consequence brought about by the loss of stability is known, IP assessment allows the instability risk to be evaluated. A risk value provides a quantitative measure for undertaking adequate preventive control actions for stability improvement and avoiding the need for conservative or “worst-case” criteria like those based on the classical deterministic analyses.

Hence, in order to effectively apply a probabilistic approach, a preliminary identification of the relevant statistical parameters has to be performed. This is a crucial step since they are potentially infinite: the choice can be made according to the required degree of accuracy.

Formally, let a proper probability space $(\mathbf{O}, P, \mathbf{S})$ be defined, where \mathbf{O} is the sample space of all possible outcomes, \mathbf{S} a sigma Algebra of events and P an additive probability measure over \mathbf{S} . For the purpose of stability investigation, the sample space \mathbf{O} may be defined as the product space $\mathbf{O} = \mathbf{O}_1 \times \mathbf{O}_2$, in which \mathbf{O}_1 is the set of all possible disturbances which can (potentially) affect system stability, \mathbf{O}_2 is the set of all possible “state vector” trajectories after the disturbance. This requires the definition of a proper “state vector” as a vector whose components are all the system variables whose values are the basis on which stability assessment is performed (see also following Eq. 2).

Let the random event I be the event of instability (over a given time horizon H of power system operation) and let (C_1, \dots, C_m) be a finite set of random events constituting all the credible—and mutually exclusive—disturbances (“contingencies”) which can affect the system operation in H and potentially make system stability worse. Then, the IP in the interval H is provided, according to the total probability theorem, by:

$$P(I) = \sum_{j=1}^m P(C_j)P(I|C_j) \quad (1)$$

where $P(C_j)$ = probability of occurrence of the disturbance C_j in the time horizon under consideration; $P(I|C_j)$ = probability of instability once the disturbance C_j occurred.

The above relation may also still hold, at least as an approximation, when the disturbances C_k are not mutually exclusive random events, provided that the joint probability of two (or more) disturbances is negligible¹: this typically happens in very short time (VST) operation which the second part of this chapter focusses on.

In the following, IP strictly denotes a term like $P(I|C)$ where C is the given fault.

As long as the fault statistics are known from available data of the system under consideration, the $P(C_j)$ terms can be considered known terms; the $P(I|C_j)$ terms

¹ Note that, without this assumption, the above equation is generally wrong, although it appears without any justification in many papers and books.

are evaluated as shown in the following so that $P(I)$ is readily obtained from the above relation.

A quite similar reasoning still applies if the fault location is modelled by an RV.

Basic RVs are load demand and the FCT. Aiming at the description of the system stability characteristics as a consequence of a given fault (for instance, a three-phase short-circuit), IP may be expressed as a function of these RVs.

For an assigned electrical power system, characterized by a state vector \mathbf{x}_0 at time t_0 in which the fault is supposed to occur, let us denote the stability region of the post-fault equilibrium point with S . Naturally, \mathbf{x}_0 is an RV (more precisely, a random vector), mainly due to the random nature of the load demand which, as previously mentioned, has a significant effect on the operation state. Let τ be the FCT: denoting by $\mathbf{x}(t, \tau; \mathbf{x}_0)$ the state vector trajectory at the time t after the fault clearing, the CCT for transient stability can be defined as follows:

$$T_{cr}(\mathbf{x}_0) = \sup\{\tau > 0 : \mathbf{x}(t, \tau, \mathbf{x}_0) \in S, \forall t > t_0 + \tau\} \tag{2}$$

This relation clearly shows the dependence of the critical time on the random initial state \mathbf{x}_0 , thus T_{cr} is also an RV. As discussed in [11–13], the FCT should also be regarded as an RV which will be denoted by T_{cl} .

By keeping in mind that the system maintains its stability conditions if and only if the FCT is smaller than the CCT, i.e. $T_{cl} < T_{cr}$, the IP for a given fault can be expressed as:

$$q = P(T_{cl} > T_{cr}) \tag{3}$$

Formally, the model in the above relation is quite similar to the “Stress–Strength” model in reliability theory: indeed, if the failure of a certain device or system is caused by the occurrence of a “stress” T_{cl} greater than the “strength” T_{cr} of the device or system, than the above probability q represents the unreliability (failure probability) of the device or system.

Once the probability distributions of T_{cl} and T_{cr} are known, $q = P(T_{cl} > T_{cr})$ may be easily computed, as is well known in probability theory, as shown below. In fact, by describing both T_{cl} and T_{cr} as continuous non-negative RVs, with joint probability density function (pdf) $f(t_{cr}, t_{cl})$, the IP q is expressed as:

$$q = P(T_{cr} < T_{cl}) = \int_{t_{cl}=0}^{t_{cl}=\infty} dt_{cl} \int_{t_{cr}=0}^{t_{cr}=t_{cl}} f(t_{cl}, t_{cr}) dt_{cr} \tag{4a}$$

In practice, the two RV are always considered in the literature as being statistically independent of each other since they are related to independent phenomena (as discussed in Sect. 2.3): under this assumption, let $f_{cl}(t)$ and $f_{cr}(t)$ be the marginal pdf of the RVs T_{cl} and T_{cr} , respectively, and let $F_{cl}(t)$ and $F_{cr}(t)$ be the correspondent cumulative probability distribution functions (cdf). The above expression may then be rewritten as follows:

$$q = \int_0^{\infty} f_{cr}(t)(1 - F_{cl}(t))dt \quad (4b)$$

Alternatively, by conditioning the instability event on the values of the clearing time T_{cl} , q can also be equivalently expressed as:

$$q = \int_0^{\infty} f_{cl}(t)F_{cr}(t)dt \quad (4c)$$

In order to evaluate the probability q , the following preliminary steps have to be taken:

- load demand and clearing time randomness have to be properly characterized in terms of distributions on the basis of realistic assumptions, by also taking the available data into account;
- the CCT distribution has to be evaluated in terms of the load distribution since there is a conceptual and analytical relationship between them (this aspect will be adequately discussed in the following sections);
- finally, the evaluation of the above integral has to be performed: often, this integration requires the use of numerical computation, but the analytical approach of this method allows it to be evaluated in a closed form.

This procedure is straightforward for the single-machine case (the so-called “one-machine infinite bus” system), as discussed in [11] since an analytical expression between T_{cr} and the load demand, based on the well-known equal area criterion, can be demonstrated. Besides, in [12], the procedure was also extended to a multi-machine system by resorting to the so-called “*Extended Equal Area Criterion*” [17]. This procedure allows difficulties arising in the evaluation of T_{cr} distribution to be overcome since T_{cr} can be analytically expressed as a function of the load demand.

2.2 Probabilistic Modelling of the CCT

In this section, the functional expression between the CCT and the load demand L at the instant of contingency is discussed. Due to the random nature of the load demand evolution and the unpredictability of the instant of fault, the load active power L has to be correctly regarded as an RV. This implies that T_{cr} is also an RV and its probability distribution may be calculated in terms of the load distribution which is generally estimated by load forecasting. In a quite general way, the load demand over time can be efficiently described through a continuous random process, $L(t)$.

With reference to a generic time instant t , the load probability cumulative distribution function is denoted by $F_L(l; t)$ and is defined over the non-negative real numbers as $F_L(l; t) = P(L(t) \leq l)$, $l \geq 0$.

On the basis of the central limit theorem, it is generally assumed that $L(t)$ can be described by a Gaussian random process [18]. The functional dependence between T_{cr} and L can be described in a compact way as $T_{cr} = g(L)$ where $g(\cdot)$ is a continuous non-negative function over the positive real axis. Moreover, it can be proved that $g(\cdot)$ is a decreasing function of the active power L . Since T_{cr} is a function of the RV L and the function $g(\cdot)$ is continuous and non-negative, the CCT is also represented by a continuous non-negative RV. Once the distribution function of L is known, in principle, the distribution function of T_{cr} can be calculated by means of well-known theorems with regard to RV transformations applied to the analytical relation. Nevertheless, since the function $g(\cdot)$ is not analytically invertible, a closed-form expression for the probability distribution function of the variable T_{cr} cannot be obtained. In these cases, the problem of the distribution evaluation is often solved using a stochastic Monte Carlo simulation.

However, in [11], an approximate method for the analytical calculation of the probability distribution function of T_{cr} has been presented. The first step for the analytical evaluation is the approximation of the true characteristic $g(L)$ with a simpler, invertible, analytical function. In particular, a log-linear model has proved to be very adequate when expressing the above characteristic for any given set of electrical parameters:

$$T_{cr} = \beta_0 \exp(-\beta_1 L) \tag{5a}$$

or:

$$\ln(T_{cr}) = a - bL \quad (a = \ln(\beta_0); \quad b = \beta_1) \tag{5b}$$

The model coefficients (β_0, β_1) are positive constants (so that: a is real, b positive), depending on the electrical parameters of the system. They can be efficiently determined by performing a linear regression of the natural logarithm of T_{cr} with regard to the load; i.e. according to the least-square method, a and b are chosen as the values minimizing the sum of the square deviations:

$$S^2(a, b) = \sum_{i=1}^n (\ln T_{cr,i} - a - bL_i)^2 \tag{6}$$

the points $(T_{cr,i}, L_i; i = 1, \dots, n)$ being chosen assuming a proper step in the interval (L_1, L_n) in which they will probably occur. For instance, a $\mu \pm 4\sigma$ interval may be chosen to represent the load values generated by a Gaussian distribution with mean μ and standard deviation σ .

On the basis of the model (see Sect. 2.1), the evaluation of the cdf of T_{cr} in terms of the probability distribution function of the load L is straightforward and, for non-negative values of the CCT, it is expressed by:

$$F_{cr}(t_{cr}) = 1 - F_L \left[\frac{a - \ln(t_{cr})}{b} \right], \quad t_{cr} \geq 0 \tag{7}$$

The above cdf is, of course, equal to zero for negative values of the argument. The above relation is quite general, i.e. independent of any particular assumption

made about the load distribution. Moreover, it can be seen that the cdf of the load $L(t)$ —i.e. the function $F_L(\cdot)$ which appears in the right-hand side of the previous equation—is dependent, of course, on time t (even if this not explicitly expressed in the previous equation); therefore, the expression of the cdf of T_{cr} also depends on time, and it is valid for any particular time instant t in which the fault occurs. The hypothesis of a Gaussian distribution, generally adopted to describe load, implies a Log-Normal distribution for the CCT and this model will be used as illustrated in the sequel. It should be stressed, however, that this distribution varies with time, although this fact may not be apparent at first sight. Indeed, time will not always be represented explicitly in the relevant equations which are often referred to at given short time intervals in which the above RV—the load, and thus the CCT—may be considered constant, and also the corresponding distribution. However, in the successive interval, this distribution is subject to changes. This should be quite clear in the framework of dynamic estimation.

2.3 Analytical Evaluation of IP: A General Methodology

As previously stated, the time interval needed for fault clearing (comprehensive of the time for the fast reclosure of the faulted line) should also be regarded as an RV, here denoted by T_{cl} . The arc extinction phenomenon, in fact, is intrinsically not deterministic; the randomness of T_{cl} may also be due to imperfect switching which can depend on the wear conditions of the poles caused by previous faults. Moreover, the (random) environmental conditions (temperature, humidity) also influence the clearing time T_{cl} . The RV T_{cl} is assumed to be continuous, non-negative and independent from the time instant of fault occurrence since the above phenomena can be considered independent of those which cause the fault. According to the definition of the CCT, it is natural to define the probability of instability after a contingency occurring at a given time instant t as:

$$q = P(T_{cr} < T_{cl}) = P\{g[L(t)] < T_{cl}\} \quad (8)$$

In (8), the relation $T_{cr} = g[L(t)]$ between the CCT and the load at (intended as “immediately before”) the instant t of the contingency is explicitly presented. In order to obtain IP over a prefixed time horizon $(0, h)$, the statistics of the random process of faults should be taken into account. This means that (8) must be integrated with the probability distribution of the number of faults in $(0, h)$ which is indeed a random process. A Poisson stochastic process [10, 19] may be generally assumed as valid for this purpose. However, a different and simpler approach is possible [11]. The stability event over $(0, h)$ may be defined as the property whereby—in the whole interval—stress T_{cl} never exceeds strength T_{cr} , which depends on time t through function g . Hence, the IP can be defined as the probability that T_{cr} is exceeded by T_{cl} for at least one time instant t in $(0, h)$: this happens if and only if, as time t varies in $(0, h)$, T_{cl} exceeds the minimum value of T_{cr} attainable in this interval.

Mathematically speaking, the IP, q , can be expressed as follows:

$$q = P\{\inf g[L(t)] < T_{cl}; 0 < t < h\} \tag{9}$$

Let the peak load value, $\Lambda = \sup (L(t); 0 < t < h)$, over the interval $(0, h)$ be introduced: since $L(t)$ is a continuous random process, Λ is a continuous RV whose probability distribution function is denoted by $F_\Lambda(\lambda)$. As previously mentioned, the function $g[L(t)]$ is continuous and decreasing versus L . Therefore, the “minimum” CCT over $(0, h)$, again denoted by T_{cr} , can be expressed as follows:

$$T_{cr} = \inf g[L(t)] = g[\sup L(t)] = g(\Lambda) \tag{10}$$

Hence, the IP is expressed by:

$$P(T_{cr} < T_{cl}) = P\{g(\Lambda) < T_{cl}\} \tag{11}$$

Hence, by keeping in mind the expression in (7), T_{cr} can be expressed in terms of the peak load Λ , so that its probability distribution function $F_{cr}(t_{cr})$ can be written as follows:

$$F_{cr}(t_{cr}) = 1 - F_\Lambda\left[\frac{a - \ln(t_{cr})}{b}\right], \quad t_{cr} \geq 0 \tag{12}$$

where the constants a and b depend on the particular system, but are indeed constant with time (unless system topology changes; this case is excluded here but can be dealt with the same methodology, once it occurs, by simply computing the new values of a and b).

The problem can then be easily solved once the peak load and the clearing time distributions are known. The distribution function and the probability density function of the clearing time T_{cl} are denoted by $F(t)$ and $f(t)$, respectively. Assuming, as reasonable, that the variables T_{cl} and T_{cr} are statistically independent, the IP in Eq. 11 can then be calculated as in Eqs. 4b or 4c. This approach, taken from [13], expresses the IP over an arbitrarily large interval—once the pdf of peak load is known—and is useful if a planning horizon is being studied. It was presented here for the sake of completeness: the application in this chapter is in fact devoted to on-line stability assessment for VST applications, related to time intervals typically lasting 1 h or less, so that in those intervals, the load L may be considered as a constant, albeit unknown (random) value so that, in practice, it will be modelled through an RV instead of a stochastic process. However, such a distinction does not affect the methodology followed in the sequel, since—as anticipated—the Gaussian distribution which will be used here is widely employed for describing the load process since it is also common practice to describe the peak load uncertainty by means of a Gaussian RV whose expected value is the forecasted peak load. For very large time horizons, the Extreme value (EV) distribution is also a natural candidate for describing the peak load [20]: this model—as well as others—can also be handled in practice with no particular problems as shown in [13], using the same methodology illustrated here.

3 Analytical IP Evaluation for Gaussian Load and Log-Normal FCT

3.1 Analytical Expression of IP

In this section, the analytical expression of the statistical parameter q —the IP of the given system under a given fault occurrence—is discussed on the basis of reasonable assumptions for the distribution functions of the RVs L (and consequently T_{cr}) and T_{cl} . For the sake of notation simplicity, the RVs T_{cr} and T_{cl} will be, respectively, named T_x and T_y .

The load L is then assumed to be a $N(\mu_L, \sigma_L)$ RV: then, letting l be a given possible load value, L is characterized by the following pdf over $(-\infty < l < +\infty)$:

$$f_L(l) = \frac{1}{\sigma_L \sqrt{2\pi}} \exp \left[\frac{-(l - \mu_L)^2}{2\sigma_L^2} \right]$$

The interval of possible values is $(-\infty < l < +\infty)$ only theoretically, being derived from the Gaussian representation. In fact, the probability of negative values for L should, of course, be equal to zero; in practice, it is known that $P(L < l) \approx 0$ if $l < \mu - 3\sigma$.

The natural logarithm of T_x is also normally distributed on the basis of the above-discussed relationship $X = \ln(T_x) = a - bL$. Hence, the distribution of T_x can be described by a Log-Normal distribution. It can be seen that the authors have shown [11–13], by means of extended numerical simulations and adequate statistical tests, that the proposed Log-Normal model for the distribution of the critical time, when the load is a Gaussian RV, is very adequate.

The Log-Normal pdf with parameters α (scale) and β (shape) is expressed by:

$$f(t; \alpha, \beta) = \frac{1}{\sqrt{2\pi t} \beta} e^{-\frac{(\ln t - \alpha)^2}{2\beta^2}} \quad t \geq 0 \quad (13)$$

and the density $f(t)$ is zero for $t < 0$.

In expression (13) α and β represent the mean value and the standard deviation of the natural logarithm of the Log-Normal variable, respectively; the mean value μ and the standard deviation σ , corresponding to Eq. 13, are:

$$\mu = e^{\alpha + \frac{\beta^2}{2}}; \quad \sigma = \mu \sqrt{\exp(\beta^2) - 1} \quad (14)$$

In this section, T_x is thus assumed to follow a Log-Normal distribution, with parameters α_x and β_x .

From relationships (14), the parameters α_x and β_x can be obtained from the statistical parameters μ_L, σ_L and the regression coefficients a and b —denoting by X the natural logarithm of the CCT T_x —are expressed by:

$$\begin{aligned} \alpha_x &= E[X] = a - b\mu_L; \\ \beta_x &= \sqrt{\text{Var}[X]} = b\sigma_L \end{aligned} \tag{15}$$

A proper probabilistic modelling for the clearing time T_y must also be introduced.

If there is a lack of experimental data, a Gaussian distribution is often assumed. However, the Gaussian distribution does not appear to be a very adequate and flexible choice, and the Log-Normal model is used instead, as in [11–13]: the Log-Normal pdf is indeed very flexible, since it can assume a large variety of shapes with positive “skewness index” which allows for a typical long “right tail” [21] whereas the Gaussian model only allows a single shape for the distribution of the clearing time, i.e. a symmetrical (bell-shaped) distribution around the mean value which is not likely to occur in real applications. The presence of a right tail in the Log-Normal density accounts for the possibility of relatively large clearing times compared with the expected value: thus, the Log-Normal assumption corresponds to a conservative approach which is appropriate when the exact distribution is unknown. The Log-Normal assumption for T_y also permits a straightforward analytical calculation of the IP, without being restrictive, since other distributions may be adopted with the same methodology as shown in [13], requiring only elementary numerical methods.

Furthermore, if the β coefficient of the Log-Normal pdf is small enough, the pdf tends to become symmetrical and may also satisfactorily approximate a Gaussian model.

In this section, the IP computation is performed under the previously discussed hypothesis that both the clearing time T_y and the minimum CCT T_x are described by Log-Normal, independent, RVs. It is, therefore, assumed that T_y has a Log-Normal distribution with parameters α_y and β_y , with density $f_{T_y}(t_y)$ expressed by (13). As a particular case, in VST applications the FCT may be considered as a known constant (as discussed in Sect. 3). With reference to the choice of α_y and β_y , they can be related to the values of μ_{T_y} (mean value of T_y) and σ_{T_y} (standard deviation of T_y), on which some information could be known in practice.

Denoting by $v_y = \frac{\sigma_{T_y}}{\mu_{T_y}}$ the coefficient of variation (CV) of T_y , the relations specifying α_y and β_y as functions of μ_{T_y} and σ_{T_y} are the following:

$$\beta_y = \sqrt{\ln(1 + v_y^2)}, \quad \alpha_y = \ln \mu_{T_y} - \frac{\beta_y^2}{2} \tag{16}$$

Different values can be considered for the parameters μ_{T_y} and v_y , in order to establish a sensitivity analysis. The IP variability versus the mean FCT μ_{T_y} is particularly interesting since such a mean clearing time is a practical measure of the reliability level of the protection system.

The determination of the probability q , when T_x and T_y are Log-Normal and independent of each other, is now considered.

First, the following auxiliary RVs are introduced:

$$X = \ln T_x, \quad Y = \ln T_y \quad (17a)$$

$$Z = X - Y \quad (17b)$$

Under the assumed hypotheses, the probability laws of the above RV X and Y are, respectively, $N(\alpha_x, \beta_x)$ and $N(\alpha_y, \beta_y)$, where, using from now on the symbol α_x instead of α_{cr} , and generally suffixes (x, y) instead of (cr, cl) :

$$\alpha_x = E[X]; \quad \alpha_y = E[Y]; \quad \beta_x = SD[X]; \quad \beta_y = SD[Y] \quad (18a)$$

According to the well-known properties of the Gaussian distribution, the variable Z , being the difference between two independent Gaussian RVs, is also Gaussian with mean value and standard deviation given by:

$$\mu_Z = E[Z] = E[X] - E[Y] = \alpha_x - \alpha_y \quad (18b)$$

$$\sigma_Z = SD[Z] = \sqrt{\text{Var}[X] + \text{Var}[Y]} = \sqrt{\beta_x^2 + \beta_y^2} \quad (18c)$$

It is opportune, although obvious, to remark that in practice μ_z is always positive ($\alpha_x > \alpha_y$) since the FCT must always be small enough when compared with the CCT for the system to possess an acceptable level of stability, namely possess a very small IP value. This—being $IP = P(Z < 0)$ —can occur only if $E[X]$ is larger than $E[Y]$: this intuitive fact will be confirmed by the computations in the following.

By introducing the standard Gaussian distribution function:

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \quad (19)$$

the IP can be easily computed as:

$$\begin{aligned} q &= P(T_x < T_y) = P[\ln(T_x) < \ln(T_y)] = P(X < Y) = P(Z < 0) = \Phi\left(-\frac{\mu_Z}{\sigma_Z}\right) \\ &= 1 - \Phi\left(\frac{\mu_Z}{\sigma_Z}\right) \end{aligned}$$

by using the well-known property: $\Phi(-x) = 1 - \Phi(x)$, valid for each real number x .

Alternatively, the “*Complementary standard Gaussian distribution function*” (CSGCDF) can be used as we have done here:

$$\Psi(x) = 1 - \Phi(x) \quad (20)$$

Since the CSGDF $\Psi(x)$ is a strictly decreasing function of x from the value $Q(-\infty) = 1$ to the value $Q(\infty) = 0$, and $Q(0) = 0.5$ (see [Appendix 1](#) for some curves), the IP can be expressed by the more compact expression:

$$q = \Psi(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) d\xi \tag{20a}$$

where

$$u = \frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}} = \frac{E(X) - E(Y)}{\sqrt{V(X) + V(Y)}} \tag{20b}$$

The above quantity u , which plays a key role in the statistical assessment of the IP, can be defined as the ‘‘SM’’ of the system (under the given fault) since the larger the value of u , the smaller the IP. Confirmation of the fact that the relation ($\alpha_x > \alpha_y$) must always be satisfied in practice (although it is not mandatory on theoretical grounds) is that, unless this happens, the IP is greater than 0.5 (if $\alpha_x = \alpha_y$, then $u = 0 \rightarrow q = 0.5$).

3.2 A Numerical Example

As a numerical example, typical values of the mean values of FCT and CCT (which will be used in the applications in the chapter) are $\mu_x = 0.145$ s and $\mu_y = 0.10$ s, respectively.

It is, therefore, assumed that the CCT and FCT follow two independent LN distributions with mean values as above; moreover, a common CV value of 0.1 is assumed for both the CCT and the FCT (i.e. $v_x = v_y = 0.10$). The following values of ($\alpha_x, \alpha_y, \beta_x, \beta_y$) correspond to these values:

$$\alpha_x = -1.9360, \quad \beta_x = 0.0998; \quad \alpha_y = -2.3076, \quad \beta_y = 0.0998$$

having used the relations already stated:

$$\beta = \sqrt{\ln(1 + v^2)}; \quad \alpha = \ln(\mu) - \frac{\beta^2}{2}$$

It can be seen that $\alpha_x > \alpha_y$, as expected. The above values β_x and β_y in the example are equal since they depend on the CV value only.

Finally, the SM value is $u = \frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}} = 2.634$ and the IP is then evaluated as $q = \Psi(u) = 0.00423$.

For high values of the SM like the one above, it is worth noting that (see [Appendix 1](#) too) the IP is very sensitive to the variations of system parameters such as the mean FCT μ_y , as can also be seen by taking the derivative of q with regard to μ_y . For instance, if the mean FCT increases from the above 0.10 to 0.11 s (a 10% increase), with the same CVs, then the IP increases to 0.0251 (a 493% increase!). The IP variation compared with both CV variations is also very high. This is just

an example of some analytical remarks, briefly discussed below, which may be useful in actual practice.

3.3 Some Final Remarks on IP Sensitivity and Its Estimation

Deferring a more detailed illustration of the IP expression to [Appendix 1](#), this section concludes by highlighting some basic facts which are easily deduced when observing the expression of q :

$$q = \Psi(u) = \Psi\left(\frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}}\right) \quad (21)$$

Since Ψ is a decreasing function of its argument, then, as intuitive, q decreases (and stability improves) as the SM increases, e.g. the mean CCT μ_x increases, or the mean FCT μ_y decreases; for given values of μ_x and μ_y , it can be verified, also analytically (see [Appendix 1](#)), that the IP also increases when the CV of the CCT and/or FCT increases.

In other words, IP increases as the uncertainty about the above times increases: this consideration has the practical implication that, if uncertainty in load values (which entails uncertainty in the CCT) and/or in FCT is neglected (i.e. their CV values are assumed as zero), the IP may be undesirably underestimated.

It can also be seen that $q(u)$ decreases very quickly towards 0—as exemplified in the above numerical example—especially when the SM u is large enough. This and other mathematical aspects of the relation between the IP and its parameters are discussed and also illustrated graphically, with some details in [Appendix 1](#), in which a sensitivity analysis of the IP is also illustrated.

The great advantage of the proposed analytical approach—compared with numerical methods or Monte Carlo simulation—consists indeed in the very easy way that this approach enables us to perform this sensitivity analysis with regard to system parameters. This is clearly a very desirable property in view of an efficient system design (i.e. with regard to the protection system: taking decisions on how to improve performance of the protection system, lowering the mean value of the FCT, or improving data acquisition in order to reduce its SD or, with regard to the network topology: trying to devise the opportune actions in order to increase the mean value of the FCT and similar actions).

It can be seen that the IP value obtained by the above methodology is only a (statistical) point estimate of the “true” IP since it is obtained from estimated values of the true parameters α_y , α_x , β_y , β_x (as far as the CCT parameters are concerned, they are “forecasted” since they are obtained on the basis of a load forecast; the FCT parameters are estimated from available field or laboratory data). The topic of estimation is discussed in a Bayesian framework in the following sections where ML estimation is also mentioned.

The problem of IP sensitivity described above and estimation are closely related. The above results, and those in [Appendix 1](#), show that, in view of the above high sensitivity of the IP to system parameters, particular attention should be paid to developing an efficient estimation of the characteristic parameters of the CCT and FCT.

4 Bayesian Statistical Inference for Transient Stability

4.1 Introduction

Bayesian inference [22–24] is becoming more and more popular as a powerful tool in all engineering applications including recent applications to power system analysis. This section and the following one, including the last ones which focus on numerical applications, are devoted to a novel methodology for the Bayesian statistical estimation, or briefly “Bayesian estimation” of the IP. In particular, here we are interested in developing a proper methodology for making inference about IP, once prior information and experimental data are available regarding the pdf of the unknown parameters of the IP, q , the transient IP.

It has been seen that the analytical expression of the IP value requires efficient statistical estimates of the true parameters $\alpha_y, \alpha_x, \beta_y, \beta_x$ to be evaluated in actual practice (e.g. the CCT parameters, as pointed out before, depend on the load parameters, which are not known, but estimated as a consequence of a load forecast). The extreme IP sensitivity in the region of the values of practical interest (i.e. those yielding IP values of the order of $1e-3$ or less) reinforces the need for an efficient estimation.

The aim of the inference is to establish both point and interval estimates of the unknown probability $q = Q(u)$ given that the parameters $(\alpha_x, \beta_x, \alpha_y, \beta_y)$ of the two LN distributions must be estimated on the basis of the available random samples $(T_{xk}: k = 1, \dots, n)$ and $(T_{yk}: k = 1, \dots, m)$.

Bayesian inference [22–25] successfully provides a coherent and effective probabilistic framework for sequentially updating estimates of model parameters as demonstrated by the ever increasing number of publications addressed to it in both theoretical and applied fields. Bayes estimation, therefore, appears to be quite adequate for on-line sequential estimation of model parameters. For well-known reasons, moreover, it is particularly efficient (compared with traditional classic estimation, based on ML methods, briefly mentioned in the final part of this section) when rare events are of interest, as is the case here. This is so true that it is currently proposed even when there are no data (see, e.g. [26] for a recent application).

The core of the Bayesian approach is the description of all uncertainties present in the problem by means of probability, and its philosophical roots lay in the

subjective meaning of probability [25]. According to such philosophy, the unknown parameters to be estimated are considered RVs, characterized by given distributions whose meaning is not a description of their “variability” (parameters are indeed considered fixed but unknown quantities) but a description of the observer’s uncertainty about their true values. Let $\omega = (\omega_1, \omega_2, \dots, \omega_n)$ be the n -dimensional vector of the parameters to be estimated. The first step in a Bayes estimation process is to introduce—in order to express the available knowledge on the parameters before observing data—a “prior” probability distribution, characterized—in the continuous case here considered—by a joint (n -dimensional) pdf over the parameter space Ω :

$$g(\omega) = g_{1,2,\dots,n}(\omega_1, \omega_2, \dots, \omega_n), \quad \omega \in \Omega \quad (22)$$

This prior pdf is often—but not always—chosen “subjectively”, which does not mean “arbitrarily”, but means “on the basis of the knowledge available to the analyst”, also using “objective” pieces of information which in most cases could not be used in classical (frequentist) statistical estimation [22–25] which does not admit the existence of a prior pdf.

Then, the data D are observed according to a formal probability model which is assumed to represent the probabilistic mechanism for some (unknown) value of ω which has generated the observed data D . This model gives rise to the “likelihood function” (LF), $L(D|\omega)$, i.e. the conditional probability of the data, given $\omega \in \Omega$. After observing the data D , all the new (updated) available knowledge is contained in the corresponding posterior distribution of ω . This is represented by a posterior joint probability density, $g(\omega|D)$, obtained from Bayes’ theorem:

$$g(\omega|D) = \frac{L(D|\omega)g(\omega)}{\int \int \dots \int_{\Omega} L(D|\omega)g(\omega)d\omega} \quad (23)$$

where the denominator is the n -fold integral over the whole parameter space Ω . Then, if a function $\tau = \tau(\omega)$ of the parameters in ω is the subject of estimation, according to the well-known “mean square error” (MSE) criterion, the best Bayes estimate “point” estimate—denoted by τ° —is given by the posterior mean of τ , given the data D . This may be obtained by well-known rules related to the expectation of a function of RV [19] by:

$$\tau^\circ = E[\tau|D] = \int \int \dots \int_{\Omega} \tau(\omega)g(\omega|D)d\omega \quad (24)$$

The particular case $\tau(\omega) = \omega_k$ —for any given k value, $k = 1, 2, \dots, n$ —yields the Bayes estimation of any single parameter ω_k ($k = 1, 2, \dots, n$).

Alternatively, by denoting the prior pdf of τ by $h(\cdot)$, i.e. the pdf induced—by a proper manipulation of the pdf $g(\omega)$ —on the space of τ values by the transformation $\tau = \tau(\omega)$, and introducing, analogously, the posterior pdf of τ , $h(\tau|D)$, the above expectation may be obtained equivalently by the following integral:

$$\tau^\circ = E[\tau|D] = \int_{\Xi} \tau \cdot h(\tau|D) d\tau \tag{25}$$

with Ξ being the space of τ values. In practice, and also in this application, it is very difficult, if not impossible, to deduce an analytical expression for the posterior pdf of τ , and the above expectation may be more easily obtained by the integral over Ω even if, in most cases, it is evaluated numerically or by means of simulation.

Unlike classical estimation, which is inherently focussed on the point estimate, here this is only a particular piece of information: indeed, Bayesian inference aims to express all the available knowledge on the parameters not by a single value but by means of the complete posterior pdf of τ , $h(\tau|D)$, denoting by $h(\cdot)$ the pdf induced—by a proper transformation of the pdf $g(\omega)$ —on the space of τ values by the transformation $\tau = \tau(\omega)$. The point estimate is only a “synthesis” of this pdf. This pdf (which would have no meaning in the classical inference since it is regarded as an unknown constant, not an RV) is the “key” information provided by Bayes estimation since it allows any probabilistic statement about the values of τ to be expressed. Typically this pdf is used to form a “Bayesian confidence interval” (BCI) or “Bayesian credible interval” of the unknown τ , defined as:

$$\text{BCI}(\tau; \pi) = (\tau_1, \tau_2) \tag{26}$$

so that $P(\tau_1 < \tau < \tau_2) = \pi$, where π is a given probability. The BCI is expressed in terms of the posterior pdf of τ as follows:

$$P(\tau_1 < \tau < \tau_2|D) = \int_{\tau_1}^{\tau_2} h(\tau|D) d\tau = \pi, \quad 0 < \pi < 1 \tag{27}$$

In practice, the above relation is generally not sufficient to find the BCI but further requirements, such as a search for the “*Highest Posterior Density*” regions [23], allow the determination of both the unknowns (τ_1, τ_2) above.

4.2 A General Methodology for Bayesian Inference on Transient IP

Let us transpose the above concepts of Bayesian inference to the estimation of the IP:

$$q = \Psi(u), \quad u = \frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}} \tag{28}$$

with u being the SM whereas $\Psi(z)$ is the CSGCDF $\Psi(z) = 1 - \Phi(z)$:

$$\Psi(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) d\xi, \quad z \in (-\infty, +\infty) \tag{29}$$

For the purpose of Bayes estimation, in the most general case, all the parameters shall be considered unknowns. Therefore, in the Bayes approach, the four parameters $(\alpha_x, \alpha_y, \beta_x, \beta_y)$ and also the SM u and the same IP, q , are regarded as a realization of RV which will be denoted by the following capital letters in the sequel:²

$$\alpha_x \rightarrow M_x; \quad \alpha_y \rightarrow M_y; \quad \beta_x^2 \rightarrow V_x; \quad \beta_y^2 \rightarrow V_y \tag{30}$$

The symbols M and V are also chosen for “mnemonic” reasons since, as mentioned below, they correspond to mean values and variances, in particular of the logarithm of the FCT (T_x) and of the logarithm of the CCT (T_y):

$$M_x = E[\ln T_x], \quad V_x = \text{Var}[\ln T_x], \quad M_y = E[\ln T_y], \quad V_y = \text{Var}[\ln T_y] \tag{31}$$

Bayes estimation, therefore, consists of assessing prior distributions to the above parameters (M_x, M_y, V_x, V_y) , and then evaluating point and interval estimates of the unknown parameter IP, described by the RV Q , function of the RV U :

$$Q = \Psi(U), \quad U = \frac{M_x - M_y}{\sqrt{V_x + V_y}} \tag{32}$$

The above relations specify the relation between the basic RV (M_x, M_y, V_x, V_y) and the IP Q :

$$Q = Q(M_x, M_y, V_x, V_y), \quad \text{with } : (M_x, M_y) \in \mathfrak{R}^2; \quad (V_x, V_y) \in \mathfrak{R}^{+2} \tag{33}$$

These relations appear to be quite complicated, in particular due to the presence of the special function $\Psi(U)$, whatever the choice of the prior pdf of the basic RV, a topic which is dealt with below. The same argument also applies, a fortiori, to the posterior pdf. In practice, it is impossible to evaluate the pdf of Q —be it prior or posterior—analytically. It can be, however, handled numerically by resorting to a reasonable “Beta approximation”, for example—introduced in a different study by Martz et al. [27] and also used by the authors in the above-mentioned (a.m.) article [16] in the framework of ML estimation. This approximation is illustrated in [Appendix 3](#) and has been shown to be very adequate, although not being the only possible approximation, since every pdf over $(0, 1)$ which can be rather smooth and flexible may be a good candidate (possible alternative choices studied by the

² From now on we use different symbols for the four parameters—when they are considered as RV—to avoid confusion with other symbols used in this section (see [Sect. 4.4](#)) and in [Appendix 2](#) where the capitals (A_x, A_y, B_x, B_y) corresponding to $(\alpha_x, \alpha_y, \beta_x, \beta_y)$ denote specific ML estimators (it is recalled that RV are denoted by capitals).

authors are also mentioned in [Appendix 2](#)). In any case, once the numerical pdf of Q is obtained, its usefulness in this application seems to consist, first of all, in establishing a proper “upper confidence bound” for IP, i.e. a q_{UP} value that makes the probability of the “desirable” event ($Q < q_{UP}$) high enough, say 0.95 or 0.99. Therefore, by denoting this high probability value with η , interest may be focussed on the determination of a q_{UP} value so that:

$$P(Q < q_{UP}) = \eta \tag{34}$$

With a sufficiently large probability value, we are, therefore, assured that the “true” IP, Q , is smaller than an “upper bound” q_{UP} , which is the Bayes counterpart of the confidence level of $100\eta\%$.

An important characteristic, perhaps the most important one, of the Bayes inference methods is the one, already mentioned, of allowing any probabilistic statement on the values under investigation, here the IP, to be expressed, e.g. in terms of the above BCIs.

This is the core of “Bayes inference”, which is something more than pure estimation, and this is also the reason behind the heading of this section (“Bayesian inference” rather than “Bayesian estimation”). Also, in many practical cases (e.g. in order to control if some prefixed requirements or standards are met by system performances), an interval estimate may be more significant than the point estimate alone.

However, also in view of the analytical or numerical difficulties mentioned above associated with the establishment of the BCI, in actual practice, there is no doubt that the typical objective of the Bayes methods is to assess the point estimate of Q . This is the topic dealt with from now on. This point estimate of Q may be evaluated after the assessment or evaluation of the:

- prior parameters’ pdf: $g_{m_x, m_y, v_x, v_y}(m_x, m_y, v_x, v_y)$, briefly denoted as $g(m_x, m_y, v_x, v_y)$;
- the LF: $L(D|m_x, m_y, v_x, v_y)$, which is given in this case by the conditional joint pdf of the observed data (FCT values, i.e. the times T_{xk} , and CCT values, i.e. times T_{yk} , recorded in the interval of interest for the IP prediction). This joint pdf is conditional to the parameters (m_x, m_y, v_x, v_y) ;
- posterior parameters’ pdf: $g_{M_x, M_y, V_x, V_y}(m_x, m_y, v_x, v_y|D)$, briefly denoted as $g(m_x, m_y, v_x, v_y|D)$, obtained by the prior pdf and the LF by means of the Bayes theorem as illustrated above.

Finally, the Bayes estimate, denoted by Q° , of the IP Q is given in principle by the four-dimensional integral:

$$Q^\circ = E[Q|D] = \int \int \int \int_{\Omega} Q(m_x, m_y, v_x, v_y) g(m_x, m_y, v_x, v_y|D) dm_x dm_y dv_x dv_y \tag{35}$$

with Ω the parameter space above specified for the four parameters (m_x, m_y, v_x, v_y) , $Q(m_x, m_y, v_x, v_y) = \Psi(u)$ (lowercase letters are used for the single determinations of the RV being studied), $u = \frac{m_x - m_y}{\sqrt{v_x + v_y}}$, and Ψ the above CSGDF.

As far as the choice of prior pdf both for the above parameters is concerned, it is well known that the most simple natural candidates are the so-called “conjugate prior pdf” [23, 24]: this means adopting Gaussian prior pdf for the mean values M_x and M_y and Inverted Gamma prior pdf for the variances V_x and V_y (see Appendix 2). These are indeed the prior pdf of mean and variances for both the normal and the Log-Normal sampling distributions [21–24].

The above integral may appear quite cumbersome, yet its evaluation can be made—at least in some cases—relatively simple, observing that the particular form of the random IP, $Q(U)$, can be reformulated in terms of the RV: $M = M_x - M_y$, $V = V_x + V_y$, $S = \sqrt{V}$. Hence $Q = \Psi(M/S)$.

If prior and posterior information on the four parameters (M_x, M_y, V_x, V_y) is recast into prior and posterior information on the difference $M = (M_x - M_y)$ and the sum $V = (V_x + V_y)$, the above integral, hence, reduces to a double integral (with respect to the pdf of M and S) which can be solved using methods related to Bayes estimation of Gaussian probabilities [23, 24].

The above transformation between the pdf of (M_x, M_y, V_x, V_y) and those of M and V may be effected by elementary RV transformations, taking advantage of the assumed s-independence between the CCT and the FCT which logically implies the s-independence between their mean values and variances. For instance, adopting conjugate Gaussian prior pdf both for M_x and M_y , assumed as s-independent, then the prior pdf for $M = (M_x - M_y)$ is again Gaussian with obvious values of the parameters; the same holds, as is well known, for the posterior pdf. As far as the variances are concerned, the same reasoning does not apply for the above-mentioned conjugate Inverted Gamma pdf which is typically adopted as the prior pdf. However, if one is able to express information directly in terms of the sum of variances $V = (V_x + V_y)$, by using an Inverted Gamma pdf for V , the classical results of Bayes estimation for the Gaussian model (mentioned in Appendix 2) may still be applied. In general, however, if other prior models are chosen, the above estimation must be carried out numerically. This poses no particular problem nowadays since specific codes and algorithms have been devised for such purposes [22–24].

A major simplification occurs in the particular case considered in the application of this contribution, i.e. in VST applications in which the FCT may be considered in practice as a known constant (as discussed in the following section).

4.3 A Simplified Method for Bayes Estimation of the IP in the Event of VST Stability Prediction

The general theoretical problem of the Bayes estimation for the IP, discussed above, will not be pursued here in view of the VST application of this contribution. In this case, indeed, an event of instability—in a very short interval lasting

typically 1 h—is very unlikely as confirmed by the typical values of the IP illustrated in the previous section and observed in actual system practice. Observing new FCT values is a very rare event.

For practical purposes, the FCT T_y can, therefore, be considered a known constant instead of an RV. This constant value is the prior estimate of the FCT before observing data since no inference can be made out of them; i.e. T_y assumes a deterministic value, $t_y = t^*$, estimated from previous experiences.

Alternatively, the FCT may be considered as an RV, with the assumed law $LN(\alpha_y, \beta_y)$, whose parameters are characterized by their prior pdf, since it is highly improbable that new data can change our information about the FCT in a VST interval (or the data are so rarely acquired that they do not change the prior pdf much).

The two cases are equivalent, as will be shown later, so that in the sequel reference will be made to the first one (i.e. a deterministic value, t^* , of the FCT T_y is assumed).

Let t be the FCT and T_x the RV describing the CCT in the given interval under investigation. Two alternative hypotheses can be assumed for the RV describing the load: (1) the load has a constant value (i.e. it is unknown, but constant in time) due to the interval shortness; (2) the load variations with time are considered—adopting a more rigorous approach—not negligible: then, reference to the peak load is made in the interval. As previously shown, the two cases are formally equivalent.

Then, the IP value in that interval is given—by using the usual transformation from an LN cdf into a Gaussian one:

$$Q = P(T_x < t) = P[(X < \ln(t)] = \Psi\left(\frac{\alpha_x - \ln(t)}{\beta_x}\right) \tag{36a}$$

always using

$$\Psi(z) = \int_z^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\xi^2}{2}\right) d\xi$$

Obviously, the above relation could also be deduced from the general one:

$$Q = \Psi\left(\frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}}\right)$$

with $\alpha_y = \ln(t)$; $\beta_y = 0$ (since T_y is deterministic, as is its logarithm Y , so that its only assumed value, $\ln(t)$, coincides with its mean value whereas its SD β_y is zero).

The consequent IP expression is, therefore, equal to:

$$\Psi\left(\frac{\alpha_x - \tau}{\beta_x}\right); \quad \tau = \ln(t) \quad (36b)$$

which can be handled like that of a Gaussian cdf, as shown in the sequel, in order to perform a Bayes point estimation of the IP.

In the framework of Bayes estimation, let us assume that the mean value α_x of $X = \ln(T_x)$ —being T_x the CCT—is an RV, denoted as M (with analogy with the previous section) whereas its SD $\beta_x = s$ —as assumed in common practice—is known.³

Let the prior information about the unknown parameter M be described by a conjugate prior normal distribution with known parameters (m_0, s_0) , i.e. $M \sim N(m_0, s_0)$ so that the prior pdf of M (bear in mind that M , the mean value of the RV $X = \ln(T_x)$ can be negative⁴):

$$g(m) = \frac{1}{s_0\sqrt{2\pi}} \exp\left[-\frac{(m - m_0)^2}{2s_0^2}\right], \quad m \in \Re \quad (37a)$$

By using results in [Appendix 2](#), the posterior pdf of M , after observing data \mathbf{X} , is again Gaussian:

$$g(m|\mathbf{X}) = \frac{1}{s_1\sqrt{2\pi}} \exp\left[-\frac{(m - m_1)^2}{2s_1^2}\right], \quad m \in \Re \quad (37b)$$

with posterior mean and variance given by:

$$m_1 = E[M|\mathbf{X}] = \frac{s^2 m + ns_0^2 M_n}{s^2 + ns_0^2} \quad (38a)$$

$$s_1^2 = \text{Var}[M|\mathbf{X}] = \frac{s_0^2 s^2}{ns_0^2 + s^2} \quad (38b)$$

where

$$M_n = (1/N) \sum_{k=1}^N X_k; \quad s = \beta = \text{SD}[X] \quad (38c)$$

being X_k a generic log-CCT value of the sample \mathbf{X} .

³ The case in which the SD should be unknown poses no problems. Indeed it can be dealt with, implying only a little computational effort, by means of well-known methods like those mentioned in [Appendix 2](#).

⁴ In the numerical examples or applications of this chapter, measuring times in seconds as done here, both X and Y have negative mean values.

The Bayes point estimate of the IP is, therefore, given by:

$$\begin{aligned}
 Q^\circ &= \int_{-\infty}^{\infty} Q(m)g(m|X)dm \\
 &= \int_{-\infty}^{\infty} \Psi\left(\frac{m-\tau}{s}\right)\frac{1}{s_1\sqrt{2\pi}}\exp\left(-\frac{(m-m_1)^2}{2s_1^2}\right)dm
 \end{aligned}
 \tag{39}$$

which, after some manipulation, after changing the variable to: $z = \frac{m-\tau}{s}$ (see [19], in chapter titled [Transmission Expansion Planning: A Methodology to Include Security Criteria and Uncertainties Using Optimization Techniques](#)), can be shown to be equal to:

$$Q^\circ = \Psi\left(\frac{m_1-\tau}{s_1}\right) = \int_z^{\infty} \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\xi^2}{2}\right)d\xi, \quad z_1 = \frac{m_1-\tau}{s_1}
 \tag{40}$$

This estimator will be used in the final numerical application related to VST stability prediction, in which, on the basis of an adequate dynamic model of the load (and the CCT), the posterior means and variances will be updated at each time step.

In these applications, typically only one datum of the CCT is observed at each step (so in the above relations $n = 1$ will be used)—i.e. at the generic k th step—the measured or forecasted load value L_k . This “data” scarcity renders the Bayes estimation more attractive, as discussed above.

Finally, let us briefly examine the second case, mentioned above, with regard to the knowledge of the pdf of the RV T_y . Let us assume that it is an RV, and not a constant as above, letting the parameters of the RV T_y , i.e. (α_y, β_y) —denoted as (α, β) in the sequel—be distributed according to their prior pdf, which remain unchanged after every interval, since no new FCT value is obtained. Let us assume, as above, that only the mean of Y , $\alpha = \alpha_y$ is unknown with a prior conjugate Gaussian distribution $N(\mu_\alpha, \sigma_\alpha)$. Consequently, the pdf of Y (conditional to $\alpha_y = \alpha$) and the pdf of α are, respectively:

$$Y|\alpha \sim N(\alpha, \beta); \quad \alpha \sim N(\mu_0, \beta_0)$$

Then, using the total probability theorem for continuous RV [19] or known results from Bayesian estimation theory [22–24] (see also [Appendix 2](#)), it can be seen that the marginal pdf is still a Gaussian pdf:

$$Y \sim N(\mu_0, \beta^*), \quad \text{with } \beta^* = \sqrt{\beta_0^2 + \beta^2}.$$

In the light of this fact, it is not difficult to show that (40) still holds with properly re-arranged values of the constants (α_x, β_x) .

4.4 A Mention of the Classical Estimation of the IP

Here, only a brief account of classical (ML) estimation of Q is given in order to compare it with the one adopted here. Some details can be found in [21, 28].

As stated in Sect. 4.1, let us assume that the following data are available: $\mathbf{X} = (X_1, \dots, X_n)$, where $X_k = \ln(T_{xk})$, $k = 1, \dots, n$: i.e. \mathbf{X} is a random sample of n elements constituted by the natural logarithms of a CCT sample, and let $\mathbf{Y} = (Y_1, \dots, Y_m)$, where $Y_k = \ln(T_{yk})$, $k = 1, \dots, m$: i.e. \mathbf{Y} is a random sample of m elements constituted by the natural logarithms of the FCT sample which can be obtained from field or laboratory data on the system protection components, with regard to the assumed kind of fault.

By referring, for easier notation, to estimated quantities with capital letters, the most widely adopted estimators (A_x, B_x, A_y, B_y) of the above four parameters—for the well-known properties of the ML estimation [21, 28]—are given, for the LN variables under study, by:

$$A_x = \left(\frac{1}{n}\right) \sum_{k=1}^n X_k; \quad B_x = \sqrt{\left(\frac{1}{n}\right) \sum_{k=1}^n (X_k - A_x)^2} \quad (41)$$

$$A_y = \left(\frac{1}{m}\right) \sum_{k=1}^m Y_k; \quad B_y = \sqrt{\left(\frac{1}{m}\right) \sum_{k=1}^m (Y_k - A_y)^2} \quad (42)$$

These estimators indeed maximize, compared with any other function of the data, the LF $L[(\mathbf{X}, \mathbf{Y}) | (\alpha_x, \alpha_y, \beta_x, \beta_y)]$.

In practice, these estimators coincide with the sample estimators of the mean values (A_x and A_y) and standard deviations (B_x and B_y) of the Normal RV $X = \ln(T_x)$ and $Y = \ln(T_y)$, and show some desirable properties such as consistency. Moreover, the log-mean estimators A_x and A_y are also unbiased estimators of α_x and α_y , respectively.

Then, the ML estimator Q^* of q is given by:

$$Q^* = \Psi \left(\frac{A_x - A_y}{\sqrt{B_x^2 + B_y^2}} \right) \quad (43)$$

In [16], the authors analysed the classical point and interval estimates of Q based on an estimator of this type whose properties are not easy to assess.

As a final remark, it should be clear that, when prior information are available, as in most engineering applications and also in this case, the Bayes estimator definitely performs better than the ML estimators. This is especially evident in on-line estimation, as will be shown later, since very few data can be collected for inference. Typically, indeed, no data are available on FCT if the fault does not occur, and this non-occurrence is of course very likely; only one datum is available on CCT, based on the forecasted load value for the time interval under investigation.

5 Dynamic Bayesian Estimation of Mean CCT and IP for VST Applications

5.1 Introduction

In this section, devoted to VST system operation, the principle of recursive Bayesian estimation is applied for a fast and efficient on-line evaluation of the mean CCT (actually, of its natural logarithms), and thus of the IP, in a dynamic framework. This evaluation exploits:

1. the above-discussed relation between the CCT and the system load L ;
2. the probabilistic knowledge of the time evolution of the load which is generally available in VST applications.

With regard to point (1), reference is made here for illustrative purposes to a single-machine system,⁵ or to a system which is reducible to it. The above discussed log-linear characteristic is therefore assumed to hold—at any given instant (for a given network topology)—between the CCT T_x and the load L :

$$T_x = \beta_0 \exp(-\beta_1 L) \tag{44}$$

with model coefficients (β_0, β_1) which are positive known constants, depending on the electrical parameters of the system. As mentioned above, they can be determined by performing a linear regression of the natural logarithms of T_x with respect to the load. This is accomplished after computing the CCT values off-line, for the given network and fault, by means of an appropriate system model based on the classical Lyapunov direct methods for transient stability analysis and sensitivity. Therefore, by denoting—as before—the natural logarithm of T_x by X and the values of X and L at a given time instant t_k by (X_k, L_k) , respectively, the following relation is assumed:

$$X_k = a - bL_k \tag{45}$$

with

$$X_k = \ln(T_{xk}); \quad a = \ln(\beta_0) (a \in \Re); \quad b = \beta_1 (b \in \Re^+)$$

This linear relation between the logarithm of the CCT (LCCT in the following) and the system load is the basis for dynamic estimation. In particular, the proposed Bayes recursive estimation uses known results in dynamic estimation—such as the Kalman filter theory—which are well established under the hypothesis that the series $\{X_k\}$ to be estimated is a Gaussian time series.

⁵ The generalization to multi-machine systems, illustrated by the authors in [12], can be accomplished without difficulties by adopting the “Extended Equal Area Criterion”.

Such hypothesis is true if the load $L(t)$ is a Gaussian process as above assumed (generalizations to other kinds of load distribution are of course possible without particular problems).

With regard to point (2), namely, the load evolution in time, an adequate load evolution model must be chosen like those adopted for VST load forecasting algorithms.

In particular, we must consider the given time instants $t_1, t_2, \dots, t_k, \dots$ of interest for VST operation (typically, the successive hours of a certain time interval in which the network topology is assumed as fixed). Then, the following “*dynamic linear model*” (DLM) [29], or “*autoregressive model*” is often satisfactorily adopted for the stochastic process ($L_k, k = 1, 2, \dots$), which is supposed to generate the load values at times t_k according to the “*system equation*”:

$$L_{k+1} = L_k + \lambda_k \quad (k = 0, 1, 2, \dots) \quad (46a)$$

in which $\{\lambda_k\}$ is a “*White Gaussian Noise*” (WGN) sequence, i.e. a set of IID Gaussian RV with mean 0, and known SD, denoted by W . This is formally expressed as:

$$\lambda_k \sim \text{WGN}(0, W) \quad (46b)$$

The sequence is “*initiated*” by a value L_0 (load value at time $t = 0$) which is (like all the L_k values) an RV, as appropriate in a Bayes framework, with known pdf representing our prior information. It is also assumed to be a Gaussian RV (with known mean μ_{L_0} and known SD σ_{L_0}), statistically independent of any finite set of the sequence $\{\lambda_k\}$

$$L_0 \sim N(\mu_{L_0}, \sigma_{L_0}) \quad (46c)$$

The above model tries to capture a reasonable “*Markovian*” dependence between the successive random values L_{k+1} and L_k in a simple way, suitable for VST applications. However, it may be extended without excessive difficulty to cover, e.g. more complex autoregressive model, such as ARIMA processes, or non-linear models [29].

Generally, the values of the load L_k are not measurable with precision but their acquisition is subject to forecasting or measurement errors (also taking into account possible time delays or even missing values in the acquisition process). The following “*observations equation*” is typically adopted for the estimation of the DLM:

$$Y_k = L_k + v_k \quad (k = 1, 2, \dots) \quad (47a)$$

where $\{v_k\}$ is another WGN sequence, with mean 0, and known SD s_v , statistically independent of the sequence $\{\lambda_k\}$ and all the other RVs in the model:

$$v_k \sim \text{WGN}(0, S) \quad (47b)$$

The above assumptions assure that both L_k and Y_k are Gaussian sequences.

Analogously to “Kalman filtering” language [29], the basic DLM equations (46a) and (47a) can be, respectively, regarded, as the “state equation” and the “measurement equation”.

In order to define a similar DLM for the LCCT values, and repeating for convenience equation (45), let us define the sequences:

$$X_k = a - bL_k, \quad Z_k = a - bY_k, \quad \zeta_k = -b\lambda_k, \quad \eta_k = -bv_k \quad (48)$$

It is easy to see that these definitions, observing that $\{\zeta_k\}$ and $\{\eta_k\}$ are still WGN sequences, allow the definition of a DLM for the sequence of the LCCT values as follows:

$$X_{k+1} = X_k + \zeta_k, \quad Z_k = X_k + \eta_k \quad (49a)$$

where (ζ_k, η_k) are, respectively, the system and measurement noise for the DLM of the LCCT.

The above assumptions for (X_0, ζ_k, η_k) are formally expressed as follows:

$$X_0 \sim N(\mu_0, \sigma_0), \quad \zeta_k \sim \text{WGN}(0, w), \quad \eta_k \sim \text{WGN}(0, s) \quad (49b)$$

The SD w of the model and the SD s of the measures, appearing in the above relations, are clearly related to the above SD (W, S) of λ_k and v_k by the following, obvious, linear relations⁶:

$$w = bW; \quad s = bS \quad (50)$$

Finally, the initial mean and SD of the LCCT sequence X_k , i.e. those of X_0 (first equation of 49b), denoted simply by (μ_0, σ_0) are obviously expressed in terms of the corresponding initial load L_0 parameters $(\mu_{L_0}, \sigma_{L_0})$ in (46c) as follows:

$$\mu_0 = a - b\mu_{L_0}; \quad \sigma_0 = b\sigma_{L_0} \quad (51)$$

It is apparent from the second equation (48) and the above hypotheses summarized in (49a) and (49b) that Z_k , the observed LCCT Z_k , being the sum of two Gaussian independent RV, X_k and ζ_k , is still a Gaussian RV whose marginal pdf is easily deducible (it is sufficient to compute its mean value and variance, as shown below). Moreover, if X_k should be known, the conditional distribution of Z_k —being ζ_k a Gaussian RV with zero mean—would be a Gaussian one with mean equal to X_k , and known SD w . Formally⁷:

$$(Z_k|X_k) \sim N(X_k, w)$$

Therefore, in the framework adopted here for the estimation process, X_k is the unknown (unobservable) mean value of the observable Gaussian RV Z_k , with

⁶ Note that if $Y = a \pm bX$, where X and Y are RV and (a, b) constants, then $\text{SD}[Y] = |b|\text{SD}[X]$ (the SD is intrinsically non-negative).

⁷ The notation $(R|S) \sim N(\alpha, \beta)$, being R and S two RV, denotes that the conditional distribution of R , given S , is $N(\alpha, \beta)$.

known SD w . In other words, interest here is focused on the estimation of the mean value of the LCCT, so that results mentioned above (and recalled in [Appendix 2](#)) related to estimation of the unknown mean value of a Gaussian RV may be adopted. For brevity, the term “LCCT” (both in the acronym form or in the complete one) will, however, still be used in the sequel instead of the more correct “mean LCCT”.

The X_k sequence is “initiated” by a value X_0 which, based on prior information for the load L_0 , is again assumed to be a Gaussian RV, as above reported, statistically independent of any finite set of the sequences $\{\xi_k\}$ and $\{\eta_k\}$.

5.2 Estimation Methodology

Once the measurements (z_1, z_2, \dots, z_k) have been assigned until time instant t_k , the optimal dynamical state estimate \hat{X}_k of the “true” state X_k at time t_k —according to the Bayesian approach to estimation—is provided by a posteriori “MSE” minimization:

$$\text{MSE} = E \left[(\hat{X}_k - X_k)^2 | z_1, z_2, \dots, z_k \right] \quad (52)$$

This can be accomplished, as will be shown, using recursive Bayesian estimation ([Appendix 2](#), see also [28]) which is substantially resumed by the following recursive relationship.

$$\hat{x}_k = E[x_k | z_1, z_2, \dots, z_k] = {}^- \hat{x}_k + G_k(z_k - {}^- \hat{x}_k) \quad (53)$$

where ${}^- \hat{x}_k$ represents the state estimate at instant t_k , before z_k knowledge, i.e. the a priori estimate at stage k , and G_k is a constant which is obtained as shown below on the basis of the above “minimum MSE” criterion. The above relation is substantially equivalent to Kalman Filter, but is obtained using the Bayes estimation process, as discussed in [29]: this method has the advantage over the classic Kalman Filter derivation of accounting for the random nature of state X and of allowing the computation of any probabilistic statement about this state. The constant G_k corresponds to the well-known “Kalman gain” [28].

The following stages to which the Bayes procedure is applied can be defined:

Stage “0”, or “a priori” Stage: “Stage 0” means the initial stage before any observation is available. Therefore, in this stage the only available information is the a priori characterization for the RV X at time instant t_0 :

$$X_0 \sim N(\mu_0, \sigma_0) \quad (54)$$

Thus, from a Bayesian point of view with quadratic “Loss function”, the initial optimal estimate is $\hat{X}_0 = \mu_0$.

Stage 1: In this stage and in following stages, according to the Bayes methodology, two kinds of information are available, before and after the

measurement—which here is the first observation z_1 —is acquired. The first (prior) information yields the prior estimation, the latter (posterior) information yields the posterior estimation.

Before the first measurement z_1 is performed, the following a priori estimation can be given:

$$X_1 = X_0 + \xi \Rightarrow X_1 \sim N(\bar{\mu}_1, \bar{\sigma}_1) \tag{55}$$

where the prior mean value and variance⁸ are determined by:

$$\begin{aligned} \bar{\mu}_1 &= E[X_1] = E[X_0] + E[\xi] = \hat{X}_0 = \mu_0 \\ \bar{\sigma}_1^2 &= \text{Var}[X_1] = \text{Var}[X_0] + \text{Var}[\xi] = \sigma_0^2 + w^2 \end{aligned} \tag{56}$$

Once the measurement z_1 is known, the aim is directed towards X_1 estimation conditional to z_1 . Denoting by z_1 the observed realization of the RV Z_1 . Z_1 is still a Gaussian RV, with conditional mean (given X_1) equal to X_1 , and SD equal to that of η_1 , i.e. s . Formally

$$Z_1 = X_1 + \eta_1 \Rightarrow (Z_1|X_1) \sim N(X_1, s) \tag{57}$$

and since $E[Z_1|X_1] = X_1$, it can be deduced that the posterior mean (i.e. the Bayes estimate) of X_1 is:

$$\hat{X}_1 = E[X_1|z_1] = \mu_0 + \frac{\bar{\sigma}_1^2}{s^2 + \bar{\sigma}_1^2}(z_1 - \mu_0) = \mu_0 + \frac{\sigma_1^2}{s^2}(z_1 - \mu_0) \tag{58}$$

where

$$\sigma_1^2 = \frac{\bar{\sigma}_1^2 s^2}{\bar{\sigma}_1^2 + s^2} = \frac{(\sigma_0^2 + w^2)s^2}{\sigma_0^2 + w^2 + s^2} \tag{59}$$

The posterior estimate is used as the prior for the next stage according to recursive Bayesian estimation, as illustrated in [Appendix 2](#). By applying this algorithm recursively, the following result at time t_k can be obtained.

Generic Stage k: By applying recursive Bayesian estimation, we can immediately verify (e.g. by induction) that the following relation—clearly appearing as the general case of (59)—holds at time instant t_k :

$$\sigma_k^2 = \frac{(\sigma_{k-1}^2 + w^2)s^2}{\sigma_{k-1}^2 + w^2 + s^2} \tag{60}$$

Similarly, the following recursive formulation for the Bayesian estimate at time t_k can be obtained:

$$\hat{X}_k = E[X_k|z_1, z_2, \dots, z_k] = \hat{X}_{k-1} + G_k(z_k - \hat{X}_{k-1}) \tag{61}$$

⁸ A prior estimate of a parameter ζ is denoted here by $\bar{\zeta}$

where

$$G_k = \frac{\sigma_k^2}{s^2} = \frac{(\sigma_{k-1}^2 + w^2)}{\sigma_{k-1}^2 + w^2 + s^2} \quad (62)$$

As is well known, this estimate exhibits the noticeable property to minimize the posterior MSE for every time instant t_k , expressed by:

$$\text{MSE} = E \left[(\hat{X}_k - X_k)^2 | z_1, z_2, \dots, z_k \right] \quad (63)$$

Of course, the recursive procedure allows knowledge of the pdf of X_k at each stage k (needless to say, unlike the “static” Bayes estimation, the posterior pdf of X changes with time), and also allows Bayes estimation of the IP. This is directly deduced using results derived in [Sect. 4.3](#): since (61) is the posterior mean of the LCCT and σ_k^2 the posterior variance, by re-arranging [Eq. 40](#), the following recursive Bayesian estimate of the IP at time t_k is obtained:

$$\hat{Q}_k = \Psi \left(\frac{\hat{X}_k - \tau}{\sigma_k} \right) \quad (64)$$

Confidence intervals, particularly the previously illustrated “upper confidence bounds” may also be computed for both the LCCT and the IP. In the following numerical application, for sake of brevity the estimation procedure is illustrated only for the LCCT sequence, which is a Gaussian one, so that its results are more easily interpretable. Moreover, the confidence interval assessment is straightforward for the LCCT sequence whereas for the IP sequence it can be computed by applying the procedure illustrated in [Appendix 3](#) at each step using the a.m. Beta approximation, since no analytical result exists. Numerical simulations results were similar as regards parameter point estimation. A numerical example of the BCI computation is still reported in [Appendix 3](#), only for the IP, being it straightforward for the LCCT.

5.3 Concluding Remark

The proposed procedure is based on a relation (CCT–Load) which can be analysed and computed off-line—for the given network—once for all, so that on-line estimation shown here does not require the solution of the system model at each iteration. This allows the time duration of the intervals in which the stability is assessed to be shortened and is favourable to reliable and efficient security assessment. A distributed version of the proposed Kalman filtering approach can be applied in the case of large power systems [30].

6 Numerical Application of the Bayes Recursive Dynamic Model

In this section, a simple numerical application—based on typical load and CCT values and simulated patterns of the load process in time—is presented in order to illustrate the on-line estimation of the LCCT in VST operation.

The load process is assumed to follow the DLM model, with ($k = 1, 2, \dots$):

$$L_{k+1} = L_k + \lambda_k \text{ and } Y_k = L_k + v_k \tag{65}$$

in which

- $\{\lambda_k\}$ is a “WGN” sequence, $\text{WGN}(0, W)$;
- $\{v_k\}$ is another WGN sequence, $\text{WGN}(0, S)$;

the two sequences are statistically independent of each other and the other RV in the model.

For simplicity of notation, the SDs in the above WGN sequences (λ_k, v_k) are, respectively, denoted as (W, S) instead of (σ_λ, σ_v) as in the previous section. The lowercase letters (w, s) will be used for the LCCT sequence.

The evolution model of the LCCT corresponding to (65) is, as already deduced:

$$X_{k+1} = X_k + \xi_k \text{ and } Z_k = X_k + \eta_k \tag{66}$$

with the already discussed basic assumptions

$$X_0 \sim N(\mu_0, \sigma_0), \quad \xi_k \sim \text{WGN}(0, w), \quad \eta_k \sim \text{WGN}(0, s) \tag{67}$$

and with the SD of the WGN sequences (ξ_k, η_k), respectively, equal to $w = bW, s = bS$.

For the sake of a numerical example, let us assume that the starting value of the system load, L_0 , measured in p.u., is a Gaussian RV with mean $\mu_{L_0} = 0.8750$ p.u. and SD $\sigma_{L_0} = 0.0417$ p.u.

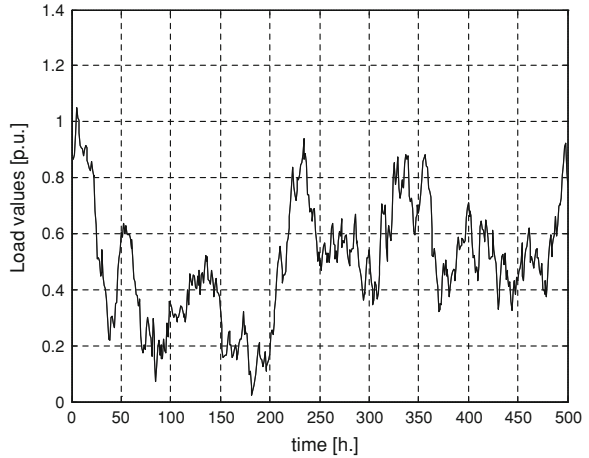
These values imply that L_0 , with probability 0.9973, assumes values in the following interval (0.75–1 p.u.) of amplitude equal to $6\sigma_{L_0}$ around the mean value μ_{L_0} .

Let us also assume that, in the log-linear model $X = a - bL$, the following values of the regression coefficients have been computed: $a = 1.7242, b = 4.1774$.

Consequently, the mean and SD of X , i.e. the parameters of the LN pdf of the CCT, are equal to $\mu_{X_0} = -1.9310$; and $\sigma_{X_0} = 0.1741$. The mean value corresponds to a CCT of about 0.145 s, which was used for the numerical examples illustrated above.

The numerical results, obtained by means of stochastic simulation of the above sequences, will be expressed in relation to the values of the “primary” SD values (S, W) of the load model and the initial load variance, V_0 , i.e. the variance of L_0 , a value which is chosen by the analyst in a Bayes methodology, on the basis of her/his prior information.

Fig. 1 Example of load pattern $\{L_k\}$, generated by a DLM (65) with parameters: $W = 0.05$, $S = 0.10$, and initial variance $V_0 = 0$



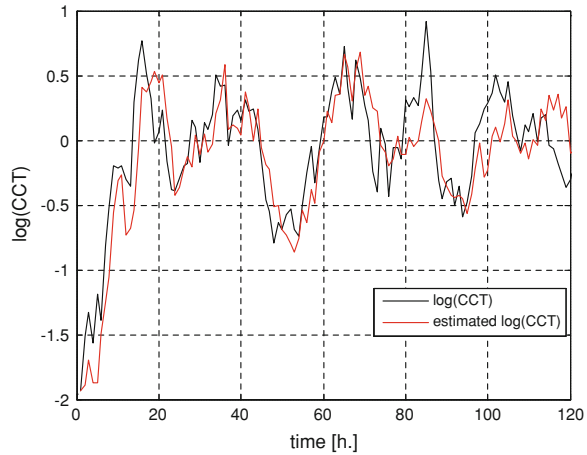
More specifically, the values (0.05 and 0.1) will be used for S and W , and these values will be swapped with each other in the course of the application to obtain at least some basic information on the sensitivity of the results in relation to variations of the model parameters. The initial variance, V_0 , of the model (i.e. the variance of X_0 , a value which is subjectively chosen by the analyst in a Bayes methodology) will alternatively also assume 2 values, 0 or 1, corresponding to different degrees of belief in the prior information (very strong in the first case, slight in the second one). An example of a possible sample path of the load sequence with these parameters (and $V_0 = 0$), simulated by means of the “*normrnd*” function of MATLAB[®], is illustrated in Fig. 1 in a time interval covering 500 h of system operation (the corresponding series of the LCCT values will be shown in Fig. 2, darker curve).

To each load sequence, generated by a DLM corresponds, as discussed above, an LCCT sequence of X_k values, also constituting a DLM, which are estimated by the recursive approach by the values X_k° . In Fig. 2, for a sample path of $N = 120$ time values, the sequence of LCCT values and of its estimated values are shown, obtained with the same values of W , S , V_0 as in Fig. 1.

The efficiency of the estimation method is evaluated using extensive Monte Carlo simulations [31]. In particular, the model performance has been summarized for any given set of time instants (t_1, t_2, \dots, t_N) by the average squared error (ASE)⁹:

⁹ The ASE index should not be confused with the MSE, which was defined in the previous section: the (theoretical) MSE evaluates the statistical mean square error between ζ_j and ζ_j^0 for any fixed time t_j with respect to the posterior conditional distribution. Instead, the ASE is an empirical measure (deduced from the sample) which takes into account the precision of estimation for *all* the RV ζ_j ($j = 1, \dots, N$) of the sequence.

Fig. 2 Example of LCCT pattern $\{X_k\}$, and its estimated values generated by a DLM as in Sect. 1.2 in chapter titled Transmission Expansion Planning: A Methodology to Include Security Criteria and Uncertainties Using Optimization Techniques, with parameters: $W = 0.05$, $S = 0.10$, and initial variance $V_0 = 0$



$$ASE = \frac{1}{N} \sum_{j=1}^N (\zeta_j^o - \zeta_j)^2 \tag{68}$$

in which ζ_j is the quantity to be estimated (in this case, the LCCT X_j at any given instant t_j) and ζ_j^o is its estimate. In practice, given the length N of a sequence (here $N = 120$ is chosen), M simulated sequences have been generated by the same algorithm and the average of the squared errors values obtained has been reported as a sample estimate of the “true” squared error. Extensive simulations were based on a number $M = 10^4$ of replications for each simulated trial; only a significant subset of the relevant results are reported in the following.

The ASE obtained using the traditional ML method has also been evaluated since the ML estimator at time k is equal—as is well known from estimation theory for the mean of a normal RV—to the sample mean of the k observed values Z_j ($j = 1, \dots, k$) so far. The precision (as measured by the relative bias and the maximum relative estimation error) of the dynamic Bayes estimator of the LCCT has also been verified. The basic statistics—estimated at the end of each simulation case study—which describe the efficiency of the proposed estimates, and which will be reported below—are:

- ASEB: average squared error of the bayes estimator;
- ASEL: average squared error of the ML estimator;
- ARE = ASEL/ASEB: average relative efficiency of the Bayes estimator compared with the ML estimator.

The ARE ratio, which is in a sense the dynamical counterpart of the classic “relative efficiency” of the Bayes estimator compared with the ML estimator used for a “static” parameter is indeed a synthetic measure of efficiency of the estimation method. The more the ARE value exceeds unity, the more efficient the Bayes estimate is when compared with the ML estimate.

A lot of different combinations of values of the model parameters (W , S , V_0) have been adopted to explore the estimation performances. In the following, the eight combinations indicated below will be reported for the triplet (W , S , V_0):

1. (0.05, 0.10, 0)
2. (0.05, 0.10, 1)
3. (0.10, 0.05, 0)
4. (0.10, 0.05, 1)
5. (0.025, 0.05, 0)
6. (0.025, 0.05, 1)
7. (0.05, 0.025, 0)
8. (0.05, 0.025, 1)

It is recalled that the above quantities are the SD describing uncertainty in the system load model (i.e. the equations in L_k and Y_k from which the ones for the LCCT are derived). The SD of the LCCT dynamic model, s and w , are larger than the correspondent load model parameters (W , S) values indicated here since $s = bS$ and $w = bW$, and $b = 4.1774$, as reported above.

It is seen that in the first four cases the values (0.05, 0.10) or (0.05, 0.10) are used for S and W , and every combination is obtained from the previous one by changing the value of the initial variance from $V_0 = 0$ to $V_0 = 1$, or by swapping the values of S and W . For instance, in cases 3 and 4, the values of W and S are swapped in relation to cases 1 and 2. In cases 1 and 3, $V_0 = 0$ was chosen; in cases 2 and 4, $V_0 = 1$ was chosen.

An analogous method has been used to form the combinations (5)–(8), by using the values (0.025, 0.05) for (W , S), i.e. half of the values (0.05, 0.10) used in the first four combinations. It is noticed that the combination of SD values of the first four combinations may be too high (particularly, because of the value 0.1 for S or W), especially for VST applications. Indeed some unrealistic value has been obtained in the course of the simulations for the LCCT (and, thus, for the IP). They have been reported here only to show that the estimation procedure works quite well even in these unrealistic cases, in which high SD values could imply high estimation errors.

Indeed, it is observed that—as typically occurs—the different choices do not affect the performances of the methodology.

Out of the many numerical simulations which have been performed, the most significant have been reported in the two tables of this section, Table 1 being relevant to the first four combinations, Table 2 relevant to the other four combinations.

For each case, the results of three different simulations (proofs), amongst all the ones performed, are reported. In particular: proof #1 is—for any given sample size—the one with the “worst” results (i.e. when the ARE gets the lowest observed value); proof #3 is the one with the “best” results (i.e. when the ARE gets the highest observed value); proof #2 gives the average results for the REFF, thus resulting intermediate between proof #1 and proof #3. So, a total number of 12 proofs is shown in each table. For example, in Table 1, the case 1.1

Table 1 Some results of the estimation performances with different combination of values of the (load) model parameters, with $(W, S) = (0.05, 0.10)$ or $(0.10, 0.05)$

Case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.05	0.10	0	0.0724	0.2495	3.4483
1.2	0.05	0.10	0	0.0606	0.4399	7.2570
1.3	0.05	0.10	0	0.0748	3.9853	53.269
2.1	0.05	0.10	1	0.0530	0.6227	11.761
2.2	0.05	0.10	1	0.0566	0.7962	14.065
2.3	0.05	0.10	1	0.0639	1.0124	15.840
3.1	0.10	0.05	0	0.0321	1.0379	32.344
3.2	0.10	0.05	0	0.0321	1.1004	34.317
3.3	0.10	0.05	0	0.0378	4.1745	110.44
4.1	0.10	0.05	1	0.0366	2.0879	57.070
4.2	0.10	0.05	1	0.0387	2.3414	60.542
4.3	0.10	0.05	1	0.0412	3.3678	81.825

V_0 assumes the values 0 or 1, here and in the following tables

(with ARE = 3.4485) precedes case 1.2 (with ARE = 7.2570). The three cases 1.1, 1.2 and 1.3 are all relevant to the same combination of values of $(W, S$ and $V_0)$ i.e. $(0.05, 0.10$ and $0)$, the first of the eight combinations above reported.

One of the results in Table 1 (the case 1.2) is also reported in Fig. 2, already mentioned.

As a general comment to the above results, it is noticed (by looking at the ASEB values) that the Bayes estimate errors are per se reasonably limited. Moreover, the relative performance in relation to the ML estimate—as measured by the ARE index—is always much greater than 1. To evaluate the precision of the estimates, other significant quantities have also been evaluated such as the average and maximum relative error of the Bayes estimates, with similar results.

It must also be remarked that, even in the “worst” cases (e.g. cases 1.1, 2.1, etc. in both tables), the ARE index is always greater than 1. Indeed, the reported results point out the efficiency of the proposed Bayesian approach, even in the case if “unrealistic” high SD as the ones in Table 1.

Finally, it has been already reported that the SD of the LCCT dynamic model, s and w , is relatively large in the application here illustrated. This could imply relatively large estimation errors whereas the reported results show that these errors are very limited, a fact which strengthens the efficiency of the estimation procedure.

For the sake of brevity, the procedure for obtaining the BCI is briefly illustrated in Appendix 3, with reference to the IP estimation.

The above good performances of the Bayes estimates with respect to the ML ones are consistent (and—to a certain extent—to be expected on theoretical grounds) with Bayesian statistical theory, as long as the Bayes estimates are evaluated assuming the “right” a priori distribution of the load, i.e. the one actually used in performing the simulation of the random samples. So, it is very

Table 2 Some results related to estimation performances with other different combination of values of the (load) model parameters with $(W, S) = (0.025, 0.05)$ or $(0.05, 0.025)$

Case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.025	0.05	0	0.0182	0.0818	4.4932
1.2	0.025	0.05	0	0.0606	0.4399	6.7520
1.3	0.025	0.05	0	0.0748	3.9853	59.619
2.1	0.025	0.05	1	0.0165	0.1736	9.280
2.2	0.025	0.05	1	0.0530	0.6227	18.167
2.3	0.025	0.05	1	0.0566	0.7962	24.506
3.1	0.05	0.025	0	0.0079	1.0096	127.52
3.2	0.05	0.025	0	0.0122	26.1034	2138.5
3.3	0.05	0.025	0	0.0134	43.40	3236.0
4.1	0.05	0.025	1	0.0090	0.1438	15.936
4.2	0.05	0.025	1	0.0102	0.4296	42.226
4.3	0.05	0.025	1	0.0089	0.9090	102.27

opportune to assess, as mentioned in the introduction, the robustness of the proposed methodology when the “a priori” hypotheses about such distribution are not valid. This is the object of the following section.

7 Some Numerical Robustness Analyses

Finally, also a simple “robustness” analysis of the proposed methodology has been performed: with respect to the initial prior distribution of the load (see Sect. 7.1), and with respect to the system load random errors prior distribution (see Sect. 7.2). For the purpose, many simulations (more than those shown here) have been carried out, assuming different (Extreme value, Log-Normal, Uniform and others) prior pdf for the initial load value L_0 or the system equation errors, instead of the Gaussian one assumed for the calculations.

7.1 A Numerical Robustness Analysis with Respect to the Initial Load Value Prior Distribution

First, some results relevant to a robustness analysis with respect to the initial load value prior distribution are reported. Since such kind of robustness is generally well established and accepted, for brevity only four cases are presented for each table, corresponding to the mean values of ASEB, ASEL, and ARE. As in the previous section, the first table is relevant to the SD values $(W, S) = (0.05, 0.10)$ or $(0.10, 0.05)$; the second table is relevant to the SD values $(W, S) = (0.025, 0.05)$ or $(0.05, 0.025)$.

Table 3 Mean values of ASEB and ASEL, and the correspondent ARE values, related to a robustness analysis with respect to a Uniform prior pdf for the initial load value. Table relevant to the SD values (W,S)=(0.05,0.10) or (0.10,0.05).

Case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1	0.05	0.10	0	0.0689	1.1970	17.380
2	0.05	0.10	1	0.0683	1.0618	15.551
3	0.10	0.05	0	0.0327	2.9910	91.572
4	0.10	0.05	1	0.0382	4.2411	111.01

In tables 3 and 4, some significant results relevant to the Uniform pdf as a prior pdf for L_0 are reported, with satisfying results which were indeed expected since the well-known robustness properties of the Kalman filter. In previous section, it was assume that the starting value of the system load, L_0 , measured in p.u., is a Gaussian RV with mean $\mu_{L_0} = 0.8750$ p.u. and SD $\sigma_{L_0} = 0.0417$ p.u. In the present case, the values of X_0 were generated, in each simulation trial, according to a Uniform prior pdf on an interval (a_L, b_L) , with the same mean and SD, so that: $a_L = 0.8028, b_L = 0.9472$.

Tables 3 and 4 report some results (the mean values of ASEB and ASEL obtained in all the simulations, and the correspondent ARE value) related to a robustness analysis of the proposed methodology with respect to a Uniform prior pdf for the initial load value L_0 instead of the Gaussian one assumed for the calculations.

The calculations were performed as if the Gaussian model, which is a basic assumption of the procedure, was the “true” model generating system errors, whilst in fact the Uniform model was the true one.

The results of this robustness analysis—and more other simulation results with different prior pdf, not shown here—still confirm the adequacy of the estimation procedure.

7.2 A Numerical Robustness Analysis with Respect to System Equation Random Errors Distribution

In addition to the previous ones, a similar robustness analysis has been performed also with respect to the random errors pdf, assumed this time to be a Uniform or an EV distribution instead of a Gaussian one, with the same mean and variance (it is reminded indeed that such parameters are assumed known). Being the pdf referred to errors distribution, they all have zero mean.

By the term “*Extreme Value*” model, it is meant the “*Largest Extreme Value*” model one characterized by the following cdf:

$$F(t; \chi, \delta) = \exp\{-\exp[-(t - \chi)/\delta]\} (-\infty < t < +\infty) \tag{69}$$

with parameters: χ real, δ positive.

Table 4 Mean values of ASEB and ASEL, and the correspondent ARE values, related to a robustness analysis with respect to a Uniform prior pdf for the initial load value: (0.025, 0.05) or (0.05, 0.025)

Case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1	0.025	0.05	0	0.0301	0.2962	9.8333
2	0.025	0.05	1	0.0158	0.9587	60.677
3	0.05	0.025	0	0.0082	0.1743	21.179
4	0.05	0.025	1	0.0068	0.8481	124.72

As already mentioned at the end of Sect. 2, the EV distribution is another natural candidate for the probabilistic description of the load, if interest is focussed on the peak load value [20]. It is indeed the most suitable for large time horizons, but it has already been shown (see Sect. 2.3) that a rigorous approach should be referred to peak load values, independently from the time horizons width. However, in this case, for the purpose of a robustness analysis, the EV model has been reported principally as an interesting alternative model since its pdf has a very different shape from the Gaussian one.

Also these results, shown in Tables 5 through 8—where they are reported in the same format of those in Sect. 6—confirm the estimation robustness. Figure 3 is referred to a typical case of those in Table 5.

All the above results show that the performances of the Bayes estimates are always scarcely sensitive to the assumed prior distribution or even to the model

Table 5 Some results related to a robustness analysis of the proposed methodology with respect to a Uniform error pdf for the system model of load sequence L_k instead of the Gaussian one assumed for the calculations

Uniform load case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.05	0.10	0	0.0697	0.3421	4.9078
1.2	0.05	0.10	0	0.0620	0.4757	7.6703
1.3	0.05	0.10	0	0.0671	0.9426	14.038
2.1	0.05	0.10	1	0.0653	0.1392	2.1324
2.2	0.05	0.10	1	0.0849	0.5900	6.9462
2.3	0.05	0.10	1	0.0664	0.7982	12.015
3.1	0.10	0.05	0	0.0386	1.2199	31.592
3.2	0.10	0.05	0	0.0344	3.0792	89.486
3.3	0.10	0.05	0	0.0406	7.2856	179.46
4.1	0.10	0.05	1	0.0312	0.6694	21.475
4.2	0.10	0.05	1	0.0386	1.2199	31.592
4.3	0.10	0.05	1	0.0406	7.2856	179.46

The different combination of values of the model parameters (W, S, V_0) are as in the tables of previous section

Table 6 Some results related to a robustness analysis of the proposed methodology with respect to a Uniform error pdf for the system model

Uniform load case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.025	0.05	0	0.0149	0.0627	4.2113
1.2	0.025	0.05	0	0.0175	0.1189	6.7963
1.3	0.025	0.05	0	0.0177	0.1952	11.017
2.1	0.025	0.05	1	0.0197	0.1218	6.1739
2.2	0.025	0.05	1	0.0202	0.2244	11.128
2.3	0.025	0.05	1	0.0137	0.5487	40.095
3.1	0.05	0.025	0	0.0089	0.4189	47.058
3.2	0.05	0.025	0	0.0091	0.8031	88.279
3.3	0.05	0.025	0	0.0114	1.1545	100.10
4.1	0.05	0.025	1	0.0109	0.5393	49.414
4.2	0.05	0.025	1	0.0106	2.4593	232.22
4.3	0.05	0.025	1	0.0094	3.4521	366.19

The table is similar to Table 5, with different values of the model parameters (W, S, V_0)

Table 7 Some results related to a robustness analysis of the proposed methodology with respect to an EV error pdf for the system model of load sequence L_k instead of the Gaussian one assumed for the calculations

EV load case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.05	0.10	0	0.0655	0.5195	7.9359
1.2	0.05	0.10	0	0.0963	1.0422	10.822
1.3	0.05	0.10	0	0.0810	2.7890	34.220
2.1	0.05	0.10	1	0.0735	0.2974	4.0455
2.2	0.05	0.10	1	0.0852	0.4999	5.8661
2.3	0.05	0.10	1	0.0594	0.3604	6.0696
3.1	0.10	0.05	0	0.0291	0.6494	22.287
3.2	0.10	0.05	0	0.0333	1.9626	58.859
3.3	0.10	0.05	0	0.0380	9.1490	241.01
4.1	0.10	0.05	1	0.0350	0.9561	27.318
4.2	0.10	0.05	1	0.0260	1.3422	51.698
4.3	0.10	0.05	1	0.0388	5.0203	129.38

The different combination of values of the model parameters (W, S, V_0) are as in the tables of previous section

distribution. In practice, the ASEB values are not significantly changed with respect to the previous cases, and it must be remarked that, even in the “worst” cases, the “efficiency” index, ARE, is always greater than 1. Similar results were obtained also for different model parameters values—i.e. different values of (W, S, V_0)—and for different random errors pdf. So, the Bayes estimator appears to be robust with respect to “wrong” prior assumptions.

Fig. 3 Curve of LCCT and its estimates, relevant to a robustness analysis with respect to a Uniform error pdf for the system model of load sequence L_k , with parameters: $W = 0.05$, $S = 0.1$, $V_0 = 0$ (1st row of Table 5)

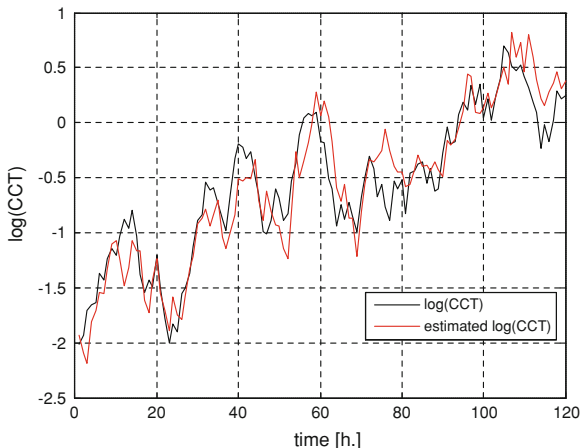


Table 8 Some results related to a robustness analysis of the proposed methodology with respect to an EV error pdf for the system

EV load case	Model parameters (W, S, V_0)			ASEB	ASEL	ARE
	W	S	V_0			
1.1	0.025	0.05	0	0.0199	0.1612	8.1102
1.2	0.025	0.05	0	0.0179	0.3679	20.533
1.3	0.025	0.05	0	0.0127	0.5091	40.193
2.1	0.025	0.05	1	0.0159	0.0781	4.9187
2.2	0.025	0.05	1	0.0198	0.1898	9.6012
2.3	0.025	0.05	1	0.0263	0.6546	24.870
3.1	0.05	0.025	0	0.0090	0.3848	42.742
3.2	0.05	0.025	0	0.0097	2.2534	231.80
3.3	0.05	0.025	0	0.0100	2.458	245.70
4.1	0.05	0.025	1	0.0068	0.2267	33.245
4.2	0.05	0.025	1	0.0116	0.9562	82.116
4.3	0.05	0.025	1	0.0110	2.6351	238.72

The table is similar to Table 5, with different values of the model parameters (W, S, V_0)

7.3 A Final Comment

The authors believe that, although deduced under simple hypotheses and/or system models, the above results—along with those already established in the field of power system or component reliability studies [32, 33]—could encourage new advanced applications of Bayesian inference in Power System analysis. Its use is indeed not yet widespread in stability or security studies, although a Bayesian classifier has been recently proposed for power system probabilistic security assessment [34]. Further developments of Bayes applications in the field of Stability surely require advanced computational tools, which are nowadays increasing in number and efficiency, as recently illustrated in [35].

Acknowledgements The authors wish to thank sincerely prof. Francesco Gagliardi, of the University of Naples Federico II, Italy, for encouraging them in undertaking the researches which constituted the foundations of this study.

8 Appendix 1: An Analytical Study of the IP

Under the assumed hypothesis of LN pdf for both the CCT and the FCT, it was shown that the IP, $q = P(\text{CCT} < \text{FCT}) = P(T_x < T_y)$, can be expressed by:

$$q = \Psi(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\zeta^2}{2}\right) d\zeta \tag{70}$$

where u is the ‘‘SM’’:

$$u = \frac{\alpha_x - \alpha_y}{\sqrt{\beta_x^2 + \beta_y^2}} = \frac{E(X) - E(Y)}{\sqrt{V(X) + V(Y)}} \tag{71}$$

being $X = \ln(T_x)$; $Y = \ln(T_y)$, and denoting by $V(X)$ and $V(Y)$ their variances.

The function $q = q(u)$ is shown in Fig. 4. Since $q(u)$ decreases very quickly towards 0, especially when u is large enough, two different curves are shown: one

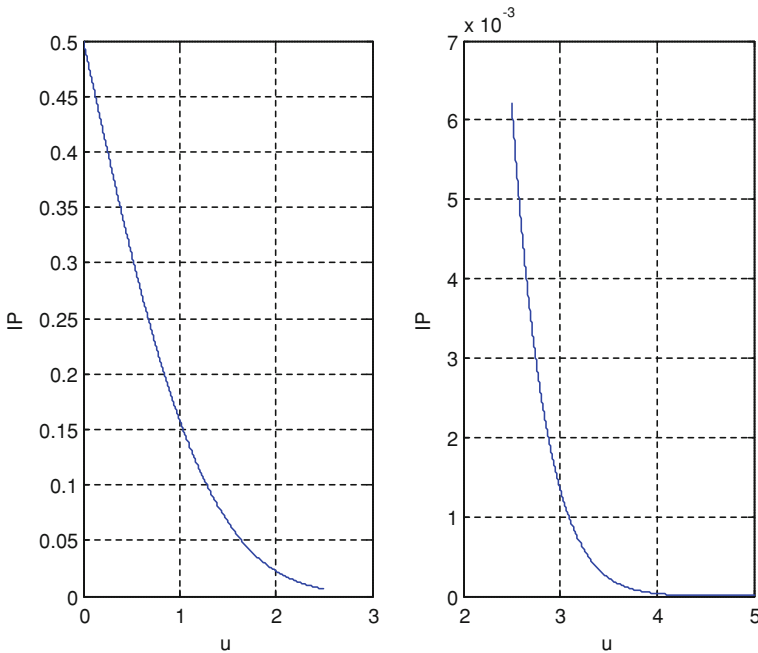


Fig. 4 A curve of the IP as a function $q = q(u)$, u being the SM

(left curve) is relevant to the interval ($0 < u < 2.5$), the other (right curve) relevant to the interval ($2.5 < u < 5$): the latter is the one which often occurs in practice since, in this interval, the IP typically assumes realistic small values, less than $6e-3$. To appreciate the quickness with which $q(u)$ decreases, the following values are given as examples:

- $q(2.0) = 2.28e-2$;
- $q(2.5) = 6.20e-3$;
- $q(3.0) = 1.30e-3$;
- $q(5.0) = 2.85e-7$

In order to appreciate the variation of the IP as a function of the SM u , the following well-known asymptotic approximation of the Ψ function which may be found in many books (e.g. [36]) is given:

$$\Psi(u) \approx \frac{1}{u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) = \frac{\phi(u)}{u}, \quad \text{for } u \text{ large enough} \quad (72)$$

with $\phi(u)$ being the standard Gaussian pdf. In practice, the above approximation is satisfactory for $u \geq 3$ (e.g. it yields 0.0015 for $u = 3$, with a relative error of less than 0.8%). From the above relation, it is readily shown (as discussed in the following) that the relative variation of q with u is in practice linear in u for typical values of u : therefore, the larger—as desired—the SM is, the more abrupt the variation (decrease) in the IP value. Indeed, a curve of the relative variation in the IP versus the argument u is shown in Fig. 5. It is apparent that such a function, which is of course negative (and decreasing), is approximately linear with u , especially for high u values.

Indeed, the relative variation of the function $Q(u)$ may be analysed using the derivative of its logarithm since:

$$\frac{dQ}{Q} = \left[\frac{Q'(u)}{Q} \right] du = D[\ln Q(u)] du. \quad (73)$$

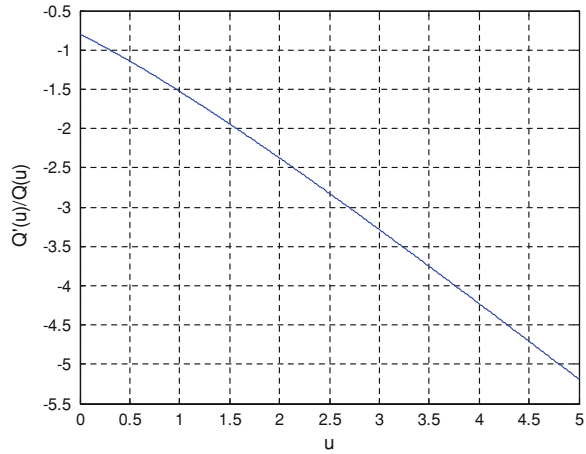
For what has been discussed above, the function: $K(u) = D[\ln(Qu)]$ tends to approach the value $(-u)$ if $u \rightarrow \infty$. However, $K(u)$ is readily expressed for any finite value of u by means of available statistical functions, since:

$$K(u) = D[\ln \Psi(u)] = -\frac{\phi(u)}{\Psi(u)} \quad (74)$$

So, the availability of the standard Gaussian pdf and cdf (e.g. the functions “*normpdf*” and “*normcdf*” in MATLAB) provides an easy computations of $K(u)$ ¹⁰ and the possibility of drawing graphs such as the one in Fig. 5.

¹⁰ In practice, the function $K(u)$ coincides with the “Hazard Rate function” of a standard Gaussian RV, as defined in Reliability applications (see, e.g. [37], where also the linearity of $h(t)$ is discussed).

Fig. 5 A curve of the relative variation of the IP as a function of the SM



Moreover, by virtue of the above asymptotic approximation, we get the above-mentioned linear approximation of $K(u)$, which is also confirmed by Fig. 5:

$$K(u) \rightarrow -u, \quad \text{as } u \rightarrow +\infty \tag{75}$$

The above relationship between the IP and the SM u can be readily expressed (see also [13]) as a function of the basic statistical parameters of the CCT—or the ones of the load on which the CCT depends—and the clearing time.

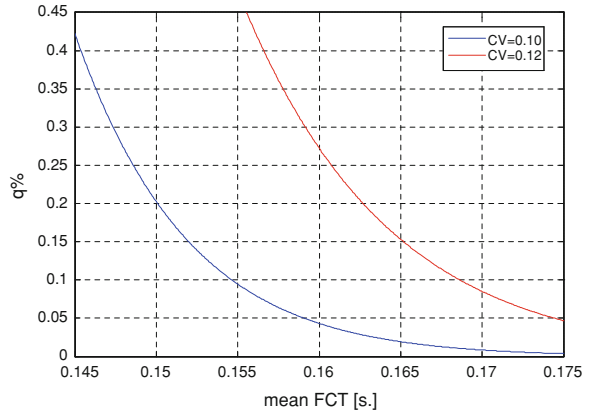
This allows a rapid sensitivity analysis of the IP for these parameters. For instance, using the already relations between the LN parameters and the mean value and the CV of the LN distributions given above, the dependence of q on μ and ν (the CV value) of both the FCT and CCT is straightforward. The following curves are obtained assuming, for illustrative purposes only, a common value ν of the CV, i.e. from the expression:

$$q = \Psi(u); \quad u = \frac{\ln \mu_x - \ln \mu_y}{\sqrt{2 \ln(1 + \nu^2)}} \tag{76}$$

The curves in Fig. 6 describe the variation of q (in %) as a function of the mean FCT, for a fixed value of the mean CCT, chosen equal to 0.1 s. as in the numerical examples of the chapter and with 2 different values of the (common) value of the CV, i.e. $\nu = 0.10$ and $\nu = 0.12$.

These curves illustrate the high or extreme variability of the IP versus the mean FCT and also the CV. This last aspect is confirmed by the curve depicted in Fig. 7 which expresses the IP—on a logarithm scale—versus the CV, assuming mean CCT value = 0.1 s and mean FCT value = 0.145 s. All the above aspects are very important in view of the estimation process, and this is why they have been illustrated in detail.

Fig. 6 Curves of the IP (in %) versus mean FCT (in s) with mean CCT = 0.1 s. Each curve refers to a given (common) value of the CV for FCT and CCT, namely CV = 0.10 (below) and 0.12 (above)



9 Appendix 2: Bayes Point Estimation for the Gaussian Model

9.1 Known Variance

For the purposes of making inference in the application in this chapter, some known results [22–24] on Bayes point estimation for the Gaussian model have been applied. Indeed, the Log-Normal model assumed for both FCT and CCT can be easily converted in the Gaussian one by means of a logarithmic transformation. Some results which are specific to the Log-Normal model are given, for example, in [21].

Let us assume that $\mathbf{X} = (X_1, \dots, X_n)$ is a random sample of n elements generated by a Gaussian model with the same mean μ , and SD σ ; let μ be an unknown to be estimated whereas σ is known. Therefore, for each $k = 1, \dots, n$, the conditional pdf of X_k , for a given value of μ is a $N(\mu, \sigma)$ pdf. Formally $X_k | \mu \sim N(\mu, \sigma)$.

Let the prior information about the unknown parameter μ be described by a prior Normal distribution with known parameters (μ_0, σ_0) , i.e. $\mu \sim N(\mu_0, \sigma_0)$,¹¹ so that the prior pdf is:

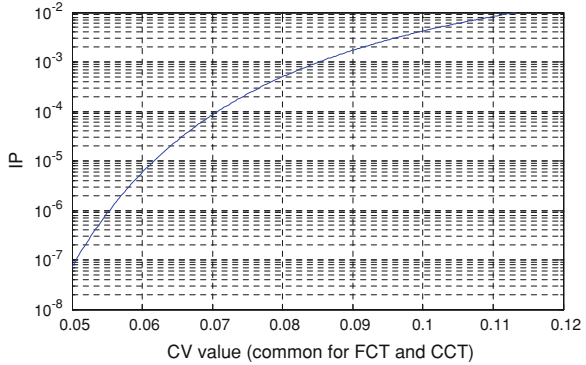
$$g(\mu) = \frac{1}{\sigma_0 \sqrt{2\pi}} \exp \left[-\frac{(\mu - \mu_0)^2}{2\sigma_0^2} \right], \quad \mu \in \Re \quad (77)$$

The prior parameters (μ_0, σ_0) are also denoted “hyper-parameters” and are assumed to be known. Before observing data, the “best” estimator of μ cannot be that its prior mean: $E[\mu] = \mu_0$, with prior variance:

$$\text{Var}[\mu] = \sigma_0^2 + \frac{\sigma^2}{n} \quad (78)$$

¹¹ The suffix “0” is typically used to denote prior parameters, e.g. (μ_0, σ_0) in this case.

Fig. 7 Curves of the IP (in %)—on a logarithm scale—versus the (supposed common) value of the CV of FCT and CCT, assuming a mean CCT value of 0.1 s and a mean FCT value of 0.145 s



The LF of the observed sample $\mathbf{X} = (X_1, \dots, X_n)$, conditional to μ , is expressed by:

$$L(\underline{X}|\mu) = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum (x_i - \mu)^2 / 2\sigma^2} \tag{79}$$

Then, multiplying the above two functions for applying the Bayes theorem, after some algebra, the following well-known result is obtained for μ° , the Bayes estimator of μ , i.e. the posterior mean:

$$\mu^\circ = E[\mu|X] = \frac{\sigma^2\mu_0 + n\sigma_0^2M_n}{\sigma^2 + n\mu_0^2} = \frac{\frac{\sigma^2}{n}\mu_0 + \sigma_0^2M_n}{\frac{\sigma^2}{n} + \sigma_0^2} \tag{80}$$

where M_n is the sample mean (which is equal in this case to the classical ML estimator of μ):

$$M_n = (1/N) \sum_{k=1}^N X_k \tag{81}$$

In the final equation expressing μ° , the prior variance (σ_0^2) and the one of M_n (σ^2/n) are clearly indicated. In this form, the above relationship shows the known property that the Bayesian estimator of μ can be, in a suggestive way, expressed as the weighted mean (a linear convex combination, in fact) of the prior estimator and the sample mean. The posterior variance is given by:

$$\text{Var}[\mu|X] = \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2} = \frac{\frac{\sigma^2}{n}\sigma_0^2}{\frac{\sigma^2}{n} + \sigma_0^2} \tag{82}$$

9.2 Unknown Variance

Although unknown variance is not considered in the application of this chapter, it seems opportune to mention it, even if very briefly.

Let us first consider known mean, $\mu = m$. The conjugate prior pdf for the variance, here denoted by V , is the so-called “*Inverted Gamma*” model, characterized by the following pdf, with argument v (a realization, of course positive, of the RV V) and (positive) parameters r and ϕ :

$$g(v; r, \phi) = \frac{\phi^r}{v^{(r+1)}\Gamma(r)} \exp(-\phi/v), \quad v > 0 \quad (83)$$

in which $\Gamma(\cdot)$ is the Euler–Gamma special function. The “*Inverted Gamma*” pdf is so denoted since it can be deduced as the pdf of the reciprocal of a Gamma RV. It is not difficult to deduce that the posterior pdf of V is again an Inverted Gamma pdf. This result derives from expressing the LF of the observed sample $\mathbf{X} = (X_1, \dots, X_n)$ as above—Eq. 79—but conditioning to $V = v$:

$$L(\underline{X}|v) = (2\pi v)^{-\frac{n}{2}} e^{-\sum (x_i - m)^2 / 2v} \quad (84)$$

(here, the mean m is assumed to be a known constant whereas the variance v is the argument under investigation). By multiplying the above prior pdf and LF, it is apparent that the posterior pdf of V is again an Inverted Gamma pdf and the updated values of r and ϕ are obvious.

Then, let us also consider the general case of unknown mean μ , i.e. both mean μ and variance V are unknown. Here, the most adopted prior model for μ is again described—conditionally to the variance V —by a Gaussian prior pdf. This prior model, multiplied by the above Inverted Gamma pdf for V , constitutes the so-called “*Normal Inverted Gamma*” prior pdf. This is indeed the conjugate prior model, as the joint posterior pdf of μ and V is again a Normal Inverted Gamma pdf [23, 24]. A similar model can also be developed in the dynamic framework [29].

10 Appendix 3: A BCI for the IP Using the Beta Distribution

In order to establish a BCI, a numerical procedure derived from a similar one, proposed in [27] and already proved satisfactory by the authors in [16], is illustrated. In [16], it was used in a different context (the one of classical statistic estimation) whereas here it is revised in the Bayesian framework. As discussed above, in the Bayesian approach the IP Q is an RV in $(0, 1)$, depending on the four random parameters $(\alpha_x, \alpha_y, \beta_x, \beta_y)$. The need for a numerical procedure is based on the fact that an analytical expression of such a pdf is impossible to find. A reasonable choice for its characterization is the approximation of its true pdf with a suitable distribution such as the Beta which is very flexible for describing RV in $(0, 1)$ and is capable of producing a large variety of shapes. The Beta is in fact the most commonly used distribution for describing random probabilities because it is also a conjugate pdf under a Binomial sampling [22–24, 36]. The analytical expression of the Beta pdf, as a function of the values q assumed by the RV Q in $(0, 1)$, is [19, 36, 37]:

$$f(q; \omega, \xi) = \begin{cases} \Gamma(\omega + \xi) \frac{q^{\omega-1} (1-q)^{\xi-1}}{\Gamma(\omega)\Gamma(\xi)} & (0 < q < 1) \\ 0 & \text{elsewhere} \end{cases} \tag{85}$$

where $\Gamma(\cdot)$ is the already introduced Gamma special function, and ω and ξ are positive shape parameters. Mean value and variance of the Beta distribution are given by:

$$\mu_B = \frac{\omega}{\omega + \xi}; \quad \sigma_B^2 = \mu_B^2 \left\{ \frac{\xi}{\omega(\omega + \xi + 1)} \right\} \tag{86}$$

In order to choose the approximating Beta pdf for Q , an adequate choice of the two parameters (ω, ξ) must be made, for instance—as proposed in [27]—by equating the above Beta statistical parameters (μ_B, σ_B^2) to opportune (as explained in the following) values of the mean value M and variance V , thus obtaining the following equations which give the Beta parameters as functions of the mean M and variance:

$$\omega = M \frac{(M - M^2 - V)}{V}; \quad \xi = \frac{\omega(1 - M)}{M} \tag{87}$$

The above mean value M and variance V of the RV Q cannot be of course obtained from its (unknown) distribution, but an excellent approximation for them is obtained: the resulting approximate values are, respectively, denoted as (M', V'). They are obtained, still following [27], by considering an expansion of $Q = \Psi(U)$ expressed as a function, say $G = G(\alpha_x, \alpha_y, \beta_x, \beta_y)$, of the four variables ($\alpha_x, \alpha_y, \beta_x, \beta_y$) in a Taylor series about the point $\Pi_0 = (A_x, A_y, B_x, B_y)$, being (A_x, A_y, B_x, B_y) the a.m. ML estimators (see Sect. 4.4) of the random parameters ($\alpha_x, \alpha_y, \beta_x, \beta_y$). In particular, expanding Q in a Taylor series about Π_0 up to second-order terms, the following values (M', V') are obtained by the well-known “Delta method” or the “statistical differentials” method [19, 36]:

$$M' = \Phi(U) - 0.5\Phi(U) \left[\frac{A_y - A_x}{B^3} \right] \chi \tag{88}$$

$$V' = \Phi(U)^2 \left[\frac{B_x^2}{nB^2} + \frac{B_y^2}{mB^2} + 0.5(A_y - A_x)^2 \frac{B_x^4}{n_1 B^6} + 0.5(A_y - A_x)^2 \frac{B_y^4}{m_1 B^6} \right] \tag{89}$$

being: $m_1 = m - 1, n_1 = n - 1$;

$$U = \frac{A_x - A_y}{\sqrt{B_x^2 + B_y^2}} \tag{90}$$

$$B = \sqrt{(B_x^2 + B_y^2)}; \tag{91}$$

$$\chi = \left[\frac{B_x^2}{n} + \frac{B_y^2}{m} + \frac{1}{2} (A_y - A_x)^2 \frac{B_x^4}{n_1 B^4} - 1.5 \frac{B_x^4}{n_1 B^2} + \frac{1}{2} (A_y - A_x)^2 \frac{B_y^4}{m_1 B^4} + 1.5 \frac{B_y^4}{m_1 B^2} \right]$$

$\Phi(x)$ and $\phi(x)$, respectively, the already introduced standard Normal cdf and pdf.

Indeed, let us denote by $F_B(q; \omega, \xi)$ the generic Beta cdf of the RV Q , evaluated in q , with parameters (ω, ξ) , i.e.

$$F_B(q; \omega, \xi) = P(Q < q) \tag{92}$$

This Beta distribution is used for inference on the BCI. For example, the estimated η -quantile of the above IP is given—still denoting by τ' the value of a true parameter τ estimated by this procedure—by:

$$Q' = F_B^{-1}(\eta; \omega', \xi') \tag{93}$$

i.e. by the inverse function of the above Beta cdf $F_B(x; \omega', \xi')$ evaluated in η , namely the solution, q^* , of: $\eta = F_B(q^*; \omega', \xi')$. So, any “upper confidence bound” for the IP mentioned in Sect. 4 (see Eq. 34) can be computed easily, since the Beta quantiles are largely available in most software packages (e.g. using the function “*Betainv*” of MATLAB). The above equation is equivalent indeed to the following:

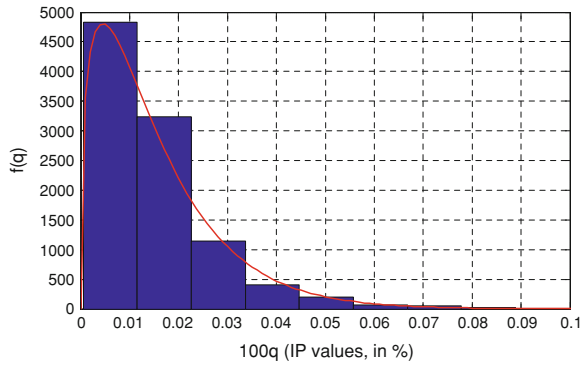
$$P(Q < Q' \eta) = \eta \tag{94}$$

And thus $Q'_\eta = F_B^{-1}(\eta; \omega', \xi')$ coincides with the upper confidence bound of probability (degree of belief) η . Of course, this procedure also allows easy estimation of the whole distribution of Q and establishes any desired confidence interval for the IP, also a bilateral one.

A simple practical numerical example is given to evaluate the BCI. In this example the FCT T_y is therefore assumed to be deterministic, with value $t_y = 0.1$ s, and the LCCT mean $\alpha_x = E[\ln(T_y)]$ is assumed to follow a prior Gaussian distribution with a mean value equal to 0.145 s and an SD equal to 1% of the mean value (i.e. a CV value equal to 0.01). The values (0.10 s, 0.145 s) of the CCT and of the mean FCT, respectively, are typical values, equal to those used in the computations already performed in the VST application of the present chapter. A CV value equal to 0.01 for the mean FCT may be also a reasonable value for describing uncertainty in such kind of on-line applications.¹² The following values of parameters $(\alpha_x, \alpha_y, \beta_x, \beta_y)$ correspond to the above CCT and FCT values:

¹² It should be remarked, however, that—in the Bayesian setting here adopted—the choice of prior parameters only reflects the information of the analyst, or her/his degree of uncertainty. So, this choice—at least from a “philosophical” point of view [25]—does not need to be “reasonable”, neither it must be necessarily accepted by others. An effort has been made nonetheless, here as in the whole chapter, to choose “realistic” values from a practical engineering point of view.

Fig. 8 Frequency histograms of the sampled IP values (in %) and the corresponding hypothetical frequency distribution obtained by the theoretical approximating Beta pdf



$\alpha_y = \ln(0.1) = -2.3026$, $\beta_y = 0$, $\beta_x = 0.0998$ whereas α_x is an RV with the above pdf.¹³

For illustrative purposes, a simulated sample of $N = 10^4$ values of $M = \alpha_x$ was generated and the corresponding empirical pdf of the IP has been evaluated and compared with the approximated theoretical Beta pdf obtained as mentioned above. The goodness of fit of this Beta pdf to the random sample has been validated through the Kolmogorov–Smirnov test of hypothesis [28]. For graphical evidence, this is also confirmed by histograms such as the one in Fig. 8 in which the frequency histograms of the sampled IP values (measured in per cent) and the corresponding hypothetical frequency distribution obtained by the a.m. Beta pdf are superimposed. Also the “ $Q-Q$ plots” [28] confirmed this adequacy, as also shown in [16].

To be more specific, the values M' and V' of the above mean and variance approximations resulted equal, for such an example, to: $M' = 0.0155\%$; $V' = (0.0155)^2$. A value of 0.0155 is therefore obtained for the SD $S' = \sqrt{V}$, with a corresponding CV equal to 0.8326, much higher than the CV = 0.1 of the basic RV M , thus confirming the already discussed high variability of the IP. This is confirmed by the values assumed by the 5th and 95th percentiles of the IP sample, i.e. 0.0032 and 0.0399, respectively, with an increase of 1147% from the former to the latter.

The Beta pdf corresponding to values of $M = M'$ and $V = V'$, which is shown in Fig. 8, has the following values of the two parameters (ω, ξ) , obtained from (M_0, V') as described above: $\omega = \omega' = 1.4047$, $\xi = \xi' = 89.2197$. This computation closes the procedure of finding a BCI.

For instance, the 0.95 upper confidence bound of the above IP is given by:

$$Q_{0.95} = F_B^{-1}(0.95; \omega', \xi') = 0.041\% \tag{95}$$

¹³ In this example, in which the mean FCT is assumed to be the only RV of the problem, the ICB could be easily computed by means of the Gaussian cdf, by using known results on Bayesian inference [23, 24, 32]. However, the presented example is kept simple on purpose, since it only serves to illustrate a methodology, which we have proven to be valid also in the general case (in which no analytical solution exists) as far as we know.

which is very close to the sample value above reported (0.0399), with a relative difference of less than 3%. So, under this approximation:

$$P(Q < 0.041\%) = 0.95 \quad (96)$$

In other words, we can be confident, with a subjective probability equal to 0.95, that the IP is less than 0.041%. As apparent, and as already discussed in Sect. 4, this information has a greater meaning than a simple point estimate, and is consistent with the establishment of possible standards.

Of course, approximately the same values of the statistical parameters of the IP and of its BCI could be obtained by performing a Monte Carlo simulation instead of computing a Beta cdf but the procedure should be repeated—in the dynamical framework here focussed on—at each time step, which is at least tedious if not time-consuming. A much more important advantage of the Beta approximation over Monte Carlo simulation is that the former allows an analytical sensitivity analysis which is cumbersome when done by simulation.

Of course, other pdf approximations can be devised for the above purposes. In our studies, an LN approximation also seemed to be adequate for describing IP randomness. However, the Beta pdf has the advantage of being theoretically limited to the interval (0, 1). Another possible adequate model in this interval is the “Negative Log-Gamma” Distribution; it was introduced in [37] and already discussed and satisfactorily adopted by the authors in studies on uncertainty characterization in reliability analyses [38], and is worth being studied also in the present context.

11 Appendix 4: Recursive Application of Bayes Estimation

Recursive Bayes estimation is based on repeated application of the Bayes theorem which shows the coherence of the updating process and its adequacy for a “dynamical” estimation, i.e. an estimation procedure involving stochastic processes. Let us perform a statistical inference for an unknown parameter θ , characterized by a prior pdf $g(\theta)$. Once a set of data D is observed, let the posterior pdf be $g(\theta|D)$:

$$g(\theta|D) = g(\theta)L(D|\theta)/P(D) \quad (97)$$

In view of dynamical applications, we can imagine acquiring data D in a two-stage process so that D consists of two sets of data—denoted by D_1 and D_2 —observed in succession. Then, by repeatedly using the Bayes theorem and the “chain rule” for joint probabilities, the above posterior pdf for θ may be alternatively obtained by expanding the previous equation in the following “two stage” process:

$$\begin{aligned} g(\theta|D_1 \cap D_2) &= g(\theta)L(D_1 \cap D_2|\theta)/P(D_1 \cap D_2) = g(\theta|D_1)L(D_2|\theta \cap D_1)/P(D_2|D_1) \\ &= g_1(\theta)L_1(D_2|\theta)/P_1(D_2) \end{aligned} \quad (98)$$

having denoted with the suffix “1” every probability (or pdf) conditional to data D_1 , i.e.

$$g_1(\theta) = g(\theta|D_1); \quad L_1(D_2|\theta) = L(D_2|\theta \cap D_1); \quad P_1(D_2) = P(D_2|D_1) \quad (99)$$

As can be seen, the posterior pdf $g(\theta|D_1 \cap D_2)$ can be obtained by applying the Bayes theorem “starting” with a prior pdf $g(\theta|D_1)$, which is the posterior pdf after observation of D_1 , then applying the same conditioning to the LF $L(D_2|\theta)$ and the probability of data D_2 . The updating process may be indefinitely continued in this way through successive stages, transforming every posterior information gained at the end of stage k into prior information for the next stage:

$$g(\theta|D_1 \cap D_2 \cdots \cap D_{k+1}) = g(\theta|D_1 \cap D_2 \cdots \cap D_k)L(D_{k+1}|\theta \cap D_1 \cdots \cap D_k)/C \quad (100)$$

where C is the “constant” (with respect to θ):

$$C = L(D_{k+1}|D_1 \cdots \cap D_k) \quad (101)$$

References

1. Kundur P (1993) Power system stability and control. Electric Power System Research, Power System Engineering Series, McGraw-Hill
2. Amjady N (2004) A framework of reliability assessment with consideration effect of transient and voltage stabilities. IEEE Trans Power Syst 19(2):1005–1014
3. Billinton R, Kuruganty PRS (1979) An approximate method for probabilistic assessment of transient stability. IEEE Trans Reliab 28(3):255–258
4. Billinton R, Kuruganty PRS (1980) A probabilistic index for transient stability. IEEE Trans PAS 99(1):195–206
5. Billinton R, Kuruganty PRS (1981) Probabilistic assessment of transient stability. IEEE Trans PAS 100(5):2163–2170
6. Billinton R, Kuruganty PRS (1981) Probabilistic assessment of transient stability in a practical multimachine system. IEEE Trans PAS 100(5):3634–3641
7. Billinton R, Kuruganty PRS (1981) Protection system modelling in a probabilistic assessment of transient stability. IEEE Trans PAS 100(7):3664–3671
8. Anderson PM, Bose A (1983) A probabilistic approach to power system stability analysis. IEEE Trans PAS 102(4):2430–2439
9. Hsu YY, Chang CL (1988) Probabilistic transient stability studies using the conditional probability approach. IEEE Trans PAS 3(4):1565–1572
10. Anders GJ (1990) Probability concepts in electric power systems. Wiley, New York
11. Chiodo E, Gagliardi F, Lauria D (1994) A probabilistic approach to transient stability evaluation. IEE Proc Generat Transm Distrib 141(5):537–544
12. Chiodo E, Lauria D (1994) Transient stability evaluation of multimachine power systems: a probabilistic approach based upon the extended equal area criterion. IEE Proc Generat Transm Distrib 141(6):545–553
13. Allella F, Chiodo E, Lauria D (2003) Analytical evaluation and robustness analysis of power system transient stability probability. Electr Eng Res Rep NR16:1–13
14. Chiodo E, Gagliardi F, La Scala M, Lauria D (1999) Probabilistic on-line transient stability analysis. IEE Proc Generat Transm Distrib 146(2):176–180

15. Ayasun S, Liang Y, Nwankpa CO (2006) A sensitivity approach for computation of the probability density function of critical clearing time and probability of stability in power system transient stability analysis. *Appl Math Comput* 176:563–576
16. Allella F, Chiodo E, Lauria D (2003) Transient stability probability assessment and statistical estimation. *Electric Power Syst Res* 67(1):21–33
17. Pavella M, Murthy PG (1994) *Transient stability of power systems. Theory and practice.* Wiley, New York
18. Breipohl AM, Lee FN (1991) A stochastic load model for use in operating reserve evaluation. In: *Proceedings of the 3rd international conference on probabilistic methods applied to electric power systems*, London, 3–5 July 1991, IEE Publishing, London
19. Papoulis A (2002) *Probability, random variables, stochastic processes.* McGraw Hill, New York
20. Belzer DB, Kellogg MA (1993) Incorporating sources of uncertainty in forecasting peak power loads. A Monte Carlo analysis using the extreme value distribution (with discussion). *IEEE Trans Power Syst* 8(2):730–737
21. Crow EL, Shimizu K (1988) *Lognormal distributions.* Marcel Dekker, New York
22. Robert CP (2001) *The Bayesian choice.* Springer Verlag, Berlin
23. Press SJ (2002) *Subjective and objective Bayesian statistics: principles, models, and applications*, 2nd edn. Wiley, New York
24. O'Hagan A (1994) *Kendall's advanced theory of statistics: vol 2b, Bayesian inference.* E. Arnold Editor, London
25. De Finetti B, Galavotti MC, Hosni H, Mura A (eds) (2008) *Philosophical lectures on probability.* Springer Verlag, Berlin
26. Jin M (2009) Estimation of reliability based on zero-failure data. In: *Proceedings of the 8th international conference on reliability, maintainability and safety, ICRMS 2009*, 20–24 July 2009, pp 308–309
27. Martz HF, Hamm LL, Reed WH, Pan PY (1993) Combining mechanistic best-estimate analysis and level 1 probabilistic risk assessment. *Reliab Eng Syst Saf* 39:89–108
28. Rohatgi VK, Saleh AK (2000) *An Introduction to Probability and Statistics*, 2nd edn. Wiley, New York
29. West M, Harrison J (1999) *Bayesian forecasting and dynamic models.* Springer Verlag, Berlin
30. Khan UA, Moura JMF (2008) Distributing the Kalman filter for large-scale systems. Part I. *IEEE Trans Signal Process* 56(10):4919–4935
31. Robert CP, Casella G (2004) *Monte Carlo statistical methods.* Springer Verlag, Berlin
32. Erto P, Giorgio M (2002) Generalised practical Bayes estimators for the reliability and the shape parameter of the Weibull distribution. In: *Proceedings of PMAAPS 2002: probabilistic methods applied to power systems*, Napoli, Italy
33. Chiodo E, Mazzanti G (2006) Bayesian reliability estimation based on a Weibull stress-strength model for aged power system components subjected to voltage surges. *IEEE Trans Dielectric Electric Insulat* 13(1):146–159
34. Kim H, Singh C (2005) Power system probabilistic security assessment using Bayes classifier. *Electric Power Syst Res* 74:157–165
35. Robert CP (2007) *Bayesian core: a practical approach to computational bayesian statistics.* Springer Verlag, Berlin
36. Johnson NL, Kotz S, Balakrishnan N (1995) *Continuous univariate distributions*, 2nd edn. Wiley, New York
37. Martz HF, Waller RA (1991) *Bayesian reliability analysis.* Krieger Publishing, Malabar, FL
38. Allella F, Chiodo E, Lauria D, Pagano M (2001) Negative log-gamma distribution for data uncertainty modelling in reliability analysis of complex systems: methodology and robustness. *Int J Qual Reliab Manag* 18(3):307–323