Sensitivity and Uncertainty Analysis in Performability Modelling

Boudewijn R. Haverkort* Adrianus M.H. Meeuwissen†

Abstract

One of the problems with performability modelling studies is the difficulty in "plugging-in" values for failure and repair rates of components. This is partly due to the lack of precise information on these parameters and partly due to the fact that such information is often regarded as confidential and therefore not published. To get an idea about the possible changes in the performability measures of interest when the input parameters change slightly, often sensitivity measures are calculated. They are however difficult to interpret and describe only local behaviour. A more global look is provided by doing an uncertainty analysis. With an uncertainty analysis, the distribution of a performability measure is derived, given (possibly dependent) distributions of the model parameters. In this paper we address these issues. We present an exact Taylor series of steady state performability measures and we discuss the validity region of this approximation. We then describe an uncertainty analysis approach based on Monte Carlo simulation. We apply both approaches to a performability model and discuss their relative merits. We vary the dependencies between model parameters and discuss their influence. For the case studies we did, the sensitivity analysis over-estimated the mean performability and under-estimated its variance.

1 Introduction

Over the last 10–15 years there has been a considerable and growing interest in the combined modelling of performance and dependability[1], so-called performability modelling, of fault-tolerant and distributed computer and communication systems [23]. This was motivated by the fact that these systems do not exhibit an *all or nothing* behaviour, so that separate performance and dependability modelling are no longer suitable.

Up till now, most of the effort in the performability modelling field has been done in the area of mathematical techniques to solve the models [23]. Less effort has been put in the aspect of conveniently describing performability models, although there is a growing interest and need for this [9]. A topic that has not been addressed in the performability modelling field is the investigation of uncertainty in the model parameters and the propagation of this uncertainty to the model outcome. We will address this issue in this paper.

From an abstract point of view, we can interpret a performability model as a function $\mathcal{F}$ of the model structure M and the model parameters $\underline{\Lambda} = (\Lambda_1, \cdots, \Lambda_n)$. The evaluation of a particular performability measure Y then boils down to evaluation of $\mathcal{F}$:

$$Y = \mathcal{F}(M, \underline{\Lambda}). \tag{1}$$

We restrict ourselves to steady state performability measures. For transient or cumulative measures, the simulation approach can be applied without modifications. The analytical approach based on sensitivity analysis however requires major changes and becomes significantly more complex. We will assume that there is no uncertainty about the structure of the performability model M. We thus have

$$Y(\underline{\Lambda}) = \mathcal{F}_M(\underline{\Lambda}). \tag{2}$$

Now, given a fixed *actual value* $\underline{\lambda}$ for $\underline{\Lambda}$ we can apply $\mathcal{F}_M$ and obtain a fixed value for $Y(\underline{\lambda}) = \mathcal{F}_M(\underline{\Lambda} := \underline{\lambda})$. In many cases however, there is uncertainty about the values $\underline{\lambda}$ that have to be "plugged in". An error not to be made is to assume that

$$E[\mathcal{F}_M(\underline{\Lambda})] = \mathcal{F}_M(E[\underline{\Lambda}]). \tag{3}$$

This is only valid when $\mathcal{F}_M$ is linear in $\underline{\Lambda}$.

When there is uncertainty about the model parameters, there are three ways to go. One can do a sensitivity analysis, an uncertainty analysis, or a perturbation analysis. We address these three approaches in Section 2. Then in Section 3, we briefly discuss

*University of Twente, Tele-Informatics and Open Systems, P.O. Box 217, 7500 AE Enschede, The Netherlands.

†Delft University of Technology, Faculty of Mathematics and Computer Science, Mekelweg 4, 2628 CD Delft, The Netherlands.

[1]We use the term dependability to signify reliability and availability. In the original definition by Laprie [19], issues concerning safety and security are also included in the concept of dependability.

1060-9857/92 $3.00 ©1992 IEEE

the class of performability models for which we derive uncertainty and sensitivity results. In Section 4 we discuss a new and exact analytical method for sensitivity analysis of performability models. Then, in Section 5 we discuss uncertainty analysis techniques based on Monte Carlo techniques. Choices that have to be made in any practical uncertainty analysis are discussed in Section 6. The problem of generating samples from dependent, multi-variate distributions is addressed in Section 7. In Section 8 we apply the described methods to a Markov reward performability model. Section 9 concludes the paper.

2 Sensitivity, uncertainty, and perturbation analysis

With *sensitivity analysis* the model $\mathcal{F}_M$ is evaluated for a specific value $\underline{\lambda}$ ("the most likely scenario"), yielding a value $Y(\underline{\lambda})$. Then, in some way, the derivative of $\mathcal{F}_M$ with respect to $\underline{\Lambda}$ is evaluated in $\underline{\Lambda} = \underline{\lambda}$, yielding insight how $Y(\underline{\lambda})$ will change when the model parameter $\underline{\lambda}$ is changed by $\underline{\epsilon}$. In no way however, this approach indicates what values for $\underline{\epsilon}$ are reasonable. Therefore, the interpretation of the derivatives alone is difficult. Still though, this approach is the only one followed up till now in the performability modelling field [1, 3, 7, 12]. The results of a sensitivity analysis can be combined with knowledge about the uncertainty of certain model parameters. We will discuss this approach in Section 4.

With *uncertainty analysis* a probability distribution $F_{\underline{\Lambda}}$ is associated with the input parameter $\underline{\Lambda}$: $F_{\underline{\Lambda}}(\underline{\lambda}) = \Pr\{\underline{\Lambda} \leq \underline{\lambda}\} = \Pr\{\Lambda_1 \leq \lambda_1, \cdots, \Lambda_n \leq \lambda_n\}$ where the random variables $\Lambda_1, \cdots, \Lambda_n$ might be dependent. When the input parameters $\underline{\Lambda}$ are random variables, Y is a random variable too. The aim is now to derive the distribution $F_Y(y) = \Pr\{Y \leq y\}$. Analytical methods to calculate $F_Y(y)$ are seldomly applicable, e.g. when the function $\mathcal{F}_M$ is linear in all its input parameters. For performability models, analytical methods can normally not be used. We therefore resort to Monte Carlo simulation. Problems arising in Monte Carlo simulation are discussed in Sections 5 through 7 [6, 13, 14, 17, 22].

With *perturbation analysis* one tries to obtain insight in the deviation $|Y(\underline{\Lambda}) - Y(\underline{\widetilde{\Lambda}})|$ of Y, given a bound on the perturbation $|\underline{\widetilde{\Lambda}} - \underline{\Lambda}|$ [8]. In the context of performability models, Van Dijk recently obtained some interesting results [5].

3 Performability modelling

Performability modelling addresses the performance as well as the dependability aspects of a system. This is normally done in the following way. A continuous time Markov chain (CTMC) $X = (X_t, t \geq 0)$ is defined on a finite state space M. The size of M is denoted $\#M$. The elements $m \in M$ are called *structure states* since every m describes the structure of the system under study in terms of the number of operational components. Given structure state m, the performance of the system is described by the reward rate $r(m)$, where $r : M \to \mathbb{R}$, i.e. per unit of time the system is in state m, a reward $r(m)$ is gained [23].

The CTMC X is completely described by the infinitisemal generator matrix $\mathbf{Q}$ and the initial probability vector $\underline{\pi}^0 = (\pi^0_1, \cdots, \pi^0_{\#M})$. The *stationary or steady state distribution* of X is denoted $\underline{\pi} = (\pi_1, \cdots, \pi_{\#\mathcal{M}})$. Whenever X is irreducible, $\underline{\pi}$ does not depend on $\underline{\pi}^0$ and is obtained by solving the system of linear equations:

$$\underline{\pi}\mathbf{Q} = \underline{0}, \quad \text{and} \quad \underline{\pi}\,\underline{1} = 1. \tag{4}$$

We can express the performability measure Y as a function of $\underline{\pi}$ and r:

$$Y = \sum_{m \in M} \pi_m r(m). \tag{5}$$

In terms of (2) we thus have $Y(\Lambda) = \mathcal{F}_M(\mathbf{Q}, r)$. We will not address uncertainty in the reward function r. Furthermore, we assume that the uncertainty in $\mathbf{Q}$ is of a parametric nature, i.e. we assume that $\mathbf{Q}$ is a function of $\underline{\Lambda}$. Consequently, we have

$$Y = \mathcal{F}_{M,r}(\mathbf{Q}(\underline{\Lambda})). \tag{6}$$

4 Sensitivity analysis

In this section we address the evaluation of the sensitivity of $Y(\underline{\Lambda})$ with respect to $\underline{\Lambda}$. We address the one dimensional case in Section 4.1. An exact Taylor series approximation for Y around Λ is discussed in Section 4.2. Finally, in Section 4.3, the sensitivity results are combined with knowledge about the uncertainty of the input parameters. In a separate paper, we will treat the multi-dimensional case [10].

4.1 1-dimensional sensitivity analysis

Let us first introduce some notation. Since we deal with only one model parameter we write Λ instead of $\underline{\Lambda}$. The k-th derivative of $Y(\Lambda)$, $\underline{\pi}(\Lambda)$ and $\mathbf{Q}(\Lambda)$ with

respect to Λ are denoted as follows:

$$Y^{(k)}(\Lambda) = \frac{\partial^k Y(\Lambda)}{\partial \Lambda^k}, \quad \underline{\pi}^{(k)}(\Lambda) = \frac{\partial^k \underline{\pi}(\Lambda)}{\partial \Lambda^k}, \quad \text{and} \quad \mathbf{Q}^{(k)}(\Lambda) = \frac{\partial^k \mathbf{Q}(\Lambda)}{\partial \Lambda^k}. \tag{7}$$

We first differentiate (4) with respect to Λ:

$$\frac{\partial}{\partial \Lambda}(\underline{\pi}(\Lambda)\mathbf{Q}(\Lambda)) = \underline{0} \Rightarrow$$
$$\underline{\pi}^{(1)}(\Lambda)\mathbf{Q}(\Lambda) + \underline{\pi}^{(0)}(\Lambda)\mathbf{Q}^{(1)}(\Lambda) = \underline{0} \Rightarrow$$
$$\underline{\pi}^{(1)}(\Lambda) = -\underline{\pi}^{(0)}(\Lambda)\mathbf{Q}^{(1)}(\Lambda)\mathbf{Q}^{-1}(\Lambda). \tag{8}$$

We can solve simultaneously for $\underline{\pi}^{(0)}(\Lambda)$ and $\underline{\pi}^{(1)}(\Lambda)$ by solving the system of linear equations:

$$(\underline{\pi}^{(0)}(\Lambda), \underline{\pi}^{(1)}(\Lambda)) \begin{pmatrix} \mathbf{Q}(\Lambda) & \mathbf{Q}^{(1)}(\Lambda) \\ \mathbf{0} & \mathbf{Q}(\Lambda) \end{pmatrix} = (\underline{0}, \underline{0}). \tag{9}$$

$Y^{(1)}(\Lambda)$ is then obtained as a linear combination of $\pi^{(1)}(\Lambda)$, by virtue of (5):

$$Y^{(1)}(\Lambda) = \sum_{m \in M} \pi_m^{(1)}(\Lambda) r(m). \tag{10}$$

The described approach can easily be generalized to involve higher derivatives. A major simplification can be obtained when we realize that $\mathbf{Q}(\Lambda)$ is generally linear in Λ. This is due to the fact that the entries of $\mathbf{Q}(\Lambda)$ signify failure rates, repair rates etc., which are generally linear combinations of the per-component failure and repair rates. This has as a consequence that $\mathbf{Q}^{(k)}(\Lambda) = \mathbf{0}$, for $k \geq 2$. Using this in the proposed generalization of (8) we obtain

$$\underline{\pi}^{(k)}(\Lambda) = -k\underline{\pi}^{(k-1)}(\Lambda)\mathbf{Q}^{(1)}(\Lambda)\mathbf{Q}^{-1}(\Lambda). \tag{11}$$

By induction to k, we can now proof for $k = 0, 1, \cdots$ that

$$\underline{\pi}^{(k)}(\Lambda) = (-1)^k k! \underline{\pi}^{(0)}(\Lambda) \left(\mathbf{Q}^{(1)}(\Lambda)\mathbf{Q}^{-1}(\Lambda)\right)^k. \tag{12}$$

4.2 A Taylor series approximation

We can use (12) to obtain the Taylor series expansion of $\underline{\pi}$ around $\Lambda = \lambda$ (the superscript (0) is omitted):

$$\underline{\pi}(\lambda + \epsilon) = \sum_{k=0}^{\infty} \underline{\pi}^{(k)}(\lambda)\frac{\epsilon^k}{k!} =$$
$$\sum_{k=0}^{\infty} (-1)^k \underline{\pi}(\lambda) \left(\epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda)\right)^k =$$
$$\underline{\pi}(\lambda) \sum_{k=0}^{\infty} \left(-\epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda)\right)^k =$$
$$\underline{\pi}(\lambda) \left(\mathbf{I} + \epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda)\right)^{-1}. \tag{13}$$

The latter equality comes from the fact that for an alternating series we have

$$\sum_{k=0}^{\infty} (-z)^k = \frac{1}{1-(-z)} = (1+z)^{-1}, \tag{14}$$

whenever the absolute values of the terms in the summation are decreasing and approaching 0. In this case this means that, for $k = 1, 2, \cdots$,

$$||(\epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda))^{k+1}|| < ||(\epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda))^k||, \tag{15}$$

and

$$\lim_{k \to \infty} ||(\epsilon \mathbf{Q}^{(1)}(\lambda)\mathbf{Q}^{-1}(\lambda))^k|| = 0 \tag{16}$$

have to hold[2]. For small models we can often derive $\underline{\pi}(\Lambda)$ symbolically. In that case we do not need the Taylor series expansian at all. For moderately sized models we can use the exact expression (13). For larger models the determination of the inverse matrices might be prohibitive. Then, only the first few elements of the infinite summation can be taken as an approximation. For an alternating series, the error made in taking only the first k elements is smaller than the absolute value of the $(k+1)$-th element.

4.3 Using sensitivities to propagate uncertainties

Now, suppose that we again deal with the vector $\underline{\Lambda}$ of uncertain model inputs. Using the methods outlined above we can numerically evaluate the partial derivatives of Y with respect to the elements of $\underline{\Lambda}$. Denote with $E[\Lambda_i]$ the first moment of Λ_i, and with $\text{var}[\Lambda_i]$ the variance of Λ_i $(i = 1, \cdots, n)$. The covariance of Λ_i and Λ_j is denoted $\text{cov}[\Lambda_i, \Lambda_j]$ $(i, j = 1, \cdots, n)$. A first order approximation of the variance $\text{var}[Y]$ in the output then equals

$$\text{var}[Y] = \sum_{i=1}^{n}\sum_{j=1}^{n} \text{cov}[\Lambda_i, \Lambda_j] \cdot \frac{\partial Y(E[\underline{\Lambda}])}{\partial \Lambda_i} \cdot \frac{\partial Y(E[\underline{\Lambda}])}{\partial \Lambda_j}. \tag{17}$$

Noting that $\text{cov}[\Lambda_i, \Lambda_i] = \text{var}[\Lambda_i]$, we can rewrite this to

$$\text{var}[Y] = \sum_{i=1}^{n} \text{var}[\Lambda_i] \cdot \left(\frac{\partial Y(E[\underline{\Lambda}])}{\partial \Lambda_i}\right)^2 +$$
$$2\sum_{i=1}^{n}\sum_{j=i+1}^{n} \text{cov}[\Lambda_i, \Lambda_j] \cdot \frac{\partial Y(E[\underline{\Lambda}])}{\partial \Lambda_i} \cdot \frac{\partial Y(E[\underline{\Lambda}])}{\partial \Lambda_j}. \tag{18}$$

[2] As a norm $||\mathbf{A}||$ of a matrix $\mathbf{A}$ one might take Euclidian norm, being the square root of the sum of the squares of the matrix entries.

Equation (18) clearly shows that dependence between the input parameters can increase as well as decrease the variance in the output measure. If the uncertain input parameters are independent, the second term in (18) vanishes.

By using (18), information about the sensitivity of performability measures with respect to model parameters is combined with knowledge about the uncertainty of these parameters. This gives better insight in the effect of parameter uncertainty than the sensitivities alone.

5 Uncertainty analysis

For many models, sensitivities are difficult or very costly to obtain. Furthermore, since sensitivity analysis only reveals local behaviour, we need other techniques to reveal more global behaviour. It is therefore that we resort to uncertainty analysis techniques based on Monte Carlo methods.

The uncertainty in the vector of parameters $\underline{\Lambda}$ is expressed by the probability distribution $F_{\underline{\Lambda}}(\underline{\lambda}) = \Pr\{\underline{\Lambda} \leq \underline{\lambda}\} = \Pr\{\Lambda_1 \leq \lambda_1, \cdots, \Lambda_n \leq \lambda_n\}$. Because $\underline{\Lambda}$ is a random variable, Y has become a random variable too. What we are interested in then, is the distribution $F_Y(y) = \Pr\{Y \leq y\}$. For most practical problems however, this distribution can not be obtained analytically, but an estimate of this distribution can often be calculated.

Suppose that we have taken m samples $\underline{\lambda}_1, \cdots, \underline{\lambda}_m$ from the distribution $F_{\underline{\Lambda}}(\underline{\lambda})$, and that we evaluate the model for these samples, yielding, for $i = 1, \cdots, m$,

$$y_i = Y(\underline{\Lambda} := \underline{\lambda}_i). \tag{19}$$

The sample mean $\hat{\mu}[y]$ and the sample standard deviation $\hat{\sigma}[y]$ of Y can now be estimated in a standard way [24].

If we want to have more information than just the first two moments, we can estimate quantiles for Y. The p-quantile Y_p is defined as $\Pr\{Y \leq Y_p\} = p$. Assume, without loss of generality that the samples for Y are ordered: $y_1 \leq y_2 \leq \cdots \leq y_m$. Sample value y_i then is an estimate for the quantile Y_p when $p = i/m$ [16, Chapter II.16].

The more confident results we want to obtain the more computations we will need to do. When trying to save computation time we can go two directions. We can apply stratified sampling techniques to generate samples that represent their distribution more accurately, such as latin hypercube sampling (LHS) [14, 22]. We can also speed-up the individual model evaluations by constructing a *response surface* for the model. The combined use of a response surface with the original model allows for the application of the *controlled variate method* which normally reduces the sample variance [22, 24].

6 Distributions & dependencies

When doing an uncertainty analysis, the modeller has to make many choices. We discuss the choices regarding distributions in Section 6.1 and the choices regarding dependencies in Section 6.2. A latent variable approach for the modelling of dependencies is discussed in Section 6.3.

6.1 Distributions

One of the tasks in modelling is the choice of model parameter values; the issue of uncertainty analysis was motivated by this.

In a sense, the task of the modeller has not become easier. Now, we not only have to "guess" average values for the model parameters but also their distributions. However, when having done so, the uncertainty analysis might reveal that the choice does not matter that much at all, or that the choice is of extreme importance. The former implies that we need not bother to much about the exact parameter values we "plug in"; the latter implies that we need to be very careful in this respect and, maybe, that we need to refine our models.

An important question to be answered is how to choose a particular uncertainty distribution from the possibly incomplete knowledge that we have. If we know the expected value μ and the variance σ^2 of an uncertain parameter, and if we know that the parameter might take values in $(-\infty, \infty)$, the most general distribution is the normal distribution $N(\mu, \sigma^2)$. With "most general distribution" we here mean the distribution $F(x)$ with the maximal entropy H (see e.g. [18, 21]), where

$$H = -\int_{-\infty}^{\infty} f(x) \ln(f(x)) dx. \tag{20}$$

Whenever we know that we deal with a positively valued parameter on $[0, \infty)$ and we know the expected value, the distribution with maximal entropy is the (negative) exponential distribution.

Whenever we know that we deal with a parameter that takes values in $[a, b]$, the uniform distribution on $[a, b]$, denoted as $U[a, b]$, is the one with the highest possible entropy.

Very often, we only have a feeling about the order of magnitude of some parameter, e.g. we know that a failure rate is "somewhere between 10^{-2} and 10^{-4}". In these cases the most appropriate distribution might be

the loguniform distribution on this interval, denoted $L[10^{-2}, 10^{-4}]$. If V is distributed $L[10^{-a}, 10^{-b}]$, and if $U =^{10} \log V$, then U is distributed $U[-a, -b]$.

6.2 Dependencies

Dependencies between model parameters generally exist for two reasons. Either they are physically induced or they stem from the fact that parameter choices are based on common knowledge or information sources. We assume that the former cause of dependencies is captured by the model itself. The latter cause should be reflected in the choice of the distribution function $F_{\underline{\Lambda}}$. These dependencies play a major role when parameters are not estimated from measurements but merely "guessed" by experts. It is especially in these cases that the uncertainty is the highest [20]. By using so-called *latent variables* these knowledge dependencies can be taken into account. We shall discuss this approach briefly in the next section.

Dependencies between random variables are expressed by *correlations* and *rank-correlations*. The correlation $\rho(X, Y)$ of two random variables is defined as

$$\rho(X, Y) = \frac{\text{cov}[X, Y]}{\sqrt{\text{var}[X]\text{var}[Y]}}, \tag{21}$$

and expresses the degree of linear relationship between X and Y. If X and Y are independent, $\rho(X, Y) = 0$; if X and Y are a linear function of each other $\rho(X, Y) = \pm 1$.

The rank-correlation $\rho_r(X, Y)$ of two random variables measures the degree of monotone relationship between X and Y, and is defined as

$$\rho_r(X, Y) = \rho(F_X(X), F_Y(Y)). \tag{22}$$

The rank-correlation is a distribution free measure of dependence between two stochastic variables and expresses the tendency of large (small) values of X coming together with large (small) values of Y.

6.3 Modelling knowledge dependence

The model described in this section tries to capture the main characteristic of knowledge dependence and to quantify its "first order" influence on the prediction results of the models under consideration. An important goal of the modelling of knowledge dependence is to make the model builder look for the best information he can find to assess the parameters in his model.

The basic idea is to couple the (random) model parameters $\underline{\Lambda}$ to *latent variables* $\underline{\mathcal{L}}$ that may be identified with the information sources. The degree of coupling depends on the "degree of subjectivity" of the information source on which the uncertainty distribution of the parameters is assessed, i.e. for parameters Λ_i whose distribution is obtained from source $\mathcal{L}_j$, more subjectivity in $\mathcal{L}_j$ means that the rank correlation $\rho_r(\Lambda_i, \mathcal{L}_j)$ is higher.

Some *arbitrariness* occurs here: information sources may be ordered according to (increasing) subjectivity, but this ordering does not imply any natural quantification for the degree of coupling. Two variables each having rank correlation ρ_r with $\mathcal{L}_j$ have an approximate rank correlation ρ_r^2 between each other (the precise value depends on the exact bivariate distribution) [20].

Conditional upon the value of $\underline{\mathcal{L}}$, the uncertainty distributions of the parameters are independent. Consequently, only two-dimensional probability distributions have to be used.

7 Sampling multivariate distributions

Using a Monte Carlo simulation technique for parameter uncertainty analysis of a model $Y(\underline{\Lambda}) = \mathcal{F}_{M,r}(\underline{\Lambda})$ basically means drawing samples $\underline{\lambda}_1, \ldots, \underline{\lambda}_m$ from the distribution $F_{\underline{\Lambda}}(\underline{\lambda}) = \Pr\{\underline{\Lambda} \leq \underline{\lambda}\} = \Pr\{\Lambda_1 \leq \lambda_1, \cdots, \Lambda_n \leq \lambda_n\}$ and calculating $\underline{y}_i = \mathcal{F}_{M,r}(\underline{\Lambda} := \underline{\lambda}_i)$ for $i = 1, 2, \ldots, m$. Thus, a sample $\{(\underline{\lambda}_1, \underline{y}_1), \ldots, (\underline{\lambda}_m, \underline{y}_m)\}$ is constructed that is analysed by common statistical methods. In this section we will be concerned with the question how to draw samples from a given multivariate distribution $F_{\underline{\Lambda}}(\underline{\lambda})$. We consider only the case where each of the samples $\underline{\lambda}_1, \ldots, \underline{\lambda}_m$ is drawn independently.

Univariate distributions. Drawing a realization λ for a one-dimensonal random variable Λ with probability distribution F_Λ can be done by the *inverse transform* technique. A $U[0, 1]$ distributed number u is drawn, e.g. by a (pseudo) random number generator, and this number is transformed by the inverse of the probability distribution of Λ:

$$\lambda := F_\Lambda^{-1}(u) \; . \tag{23}$$

F is always invertible by defining $F_\Lambda^{-1}(u)$ to equal the minimum value λ for which $F_\Lambda(\lambda) \geq u$. For particular distribution functions, e.g. Gaussian, special algorithms exist for the generation of samples [15].

Independent multivariate distributions. Often independence is assumed between the n random variables $\Lambda_1, \ldots, \Lambda_n$. Probably not always because this assumption is so plausible, but because it makes the calculations so easy. In the case of independent random variables $\Lambda_1, \ldots, \Lambda_n$ namely, one can draw n independent $U[0, 1]$ numbers $u_1, \ldots, u_n$, and then apply

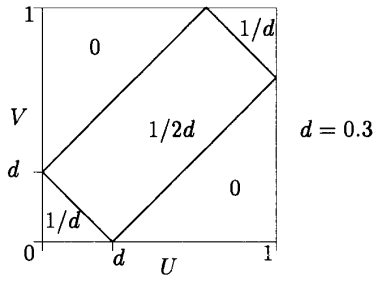

Figure 1: Diagonal band distribution $G_{0.3}(U, V)$

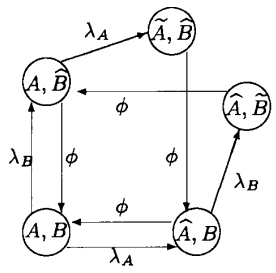

Figure 2: CTMC of a one-unit, one-standby system

the inverse transform $F_{\Lambda_i}^{-1}$ on all u_i $(i = 1, \ldots, n)$ independently, because the joint distribution in the independent case equals

$$F_{\underline{\Lambda}}(\underline{\lambda}) = \prod_{i=1}^{n} F_{\Lambda_i}(\lambda_i) \; . \tag{24}$$

Dependent multivariate distributions. In our models the marginal distributions of the random variables $\underline{\Lambda}$ will be specified, and dependence is introduced by the coupling of the observable random variables $\underline{\Lambda}$ to latent variables $\underline{\mathcal{L}}$. This coupling will be specified by a rank correlation $\rho_r(\Lambda_i, \mathcal{L}_j)$ for the observable variable Λ_i and the latent variable $\mathcal{L}_j$ (for each $i = 1, \ldots, n$, at most one rank correlation is specified).

The basic idea is to generate *dependent* $U[0, 1]$ numbers $u_1, \ldots, u_n$ and then apply the inverse transforms for each of the variables individually. This can be done in many ways, because a joint distribution is not entirely characterized by its marginals and correlations. For instance one could use dependent Gaussian variables for the generation of dependent uniform ones [15, 17]. Alternatively, interesting from a theoretical point of view, one might use that joint distribution for the generation of dependent uniforms which has maximal entropy given the specified correlations [21]. We used for our calculations the computationally very attractive so-called *diagonal band distributions* [4].

Diagonal band distribution. The diagonal band distribution $G_d(U, V)$ is a one-parameter bivariate distribution defined on the unit square $[0, 1] \times [0, 1]$. Our discussion relates to the parameter range $0 \leq d \leq 1$ which involves the positive correlations. For negative correlation values an analogues discussion applies. Between parameter d and the correlation $\rho(U, V)$ the following relation exists [4]:

$$\rho(U, V) = d^3 - 2d^2 + 1 \; . \tag{25}$$

The density of G_d takes values 0, $1/2d$ and $1/d$ on the five regions it divides the unit square in. The diagonal band distribution is designed such that the conditional distribution $G_d(V|U = u)$ is $U[u-d, u+d]$. Obviously, for small or large values of u, the restriction of the support of V to the unit interval implies the doubled values for the density in the edges of Figure 1.

8 An application

In this section we discuss an uncertainty and a sensitivity analysis of a performability model of a one-unit, one-standby system. In the subsections to follow, we first discuss the system to be modelled, then the performability model itself. After that we discuss the analytical sensitivity analysis and the uncertainty analysis based on Monte Carlo simulation. We finally evaluate the two approaches.

8.1 The system

We consider a single processing unit A with standby B. When operational, processing unit A can fail with intensity λ_A, processing unit B with intensity λ_B. Failed units are repaired with intensity ϕ. Repair activities are scheduled in a FCFS manner. This means that whenever processing unit A fails when B is already in repair, the repair of B is completed before the repair of A starts. It is assumed that when both processing units are operational they cooperate on the completion of tasks. When one of the processing units is down, the other takes over all the work. When both processing units are down, no work is done.

8.2 The model

Assuming all failure and repair time distributions to be exponential, we can model the system as a 5-state CTMC as depicted in Figure 2. The state labels indicate whether the units are operational, down and

	Λ_A	Λ_B	ρ_r	Φ
$S1$	$L[10^{-5}, 10^{-3}]$	$L[10^{-4}, 10^{-2}]$	0.0	$U[0.5, 1.5]$
$S2$	$L[10^{-5}, 10^{-3}]$	$L[10^{-4}, 10^{-2}]$	0.5	$U[0.5, 1.5]$
$S3$	$L[10^{-5}, 10^{-3}]$	$L[10^{-4}, 10^{-2}]$	0.9	$U[0.5, 1.5]$
$S4$	$L[10^{-4}, 10^{-1}]$	$L[10^{-4}, 10^{-1}]$	0.0	$U[0.5, 1.5]$
$S5$	$L[10^{-4}, 10^{-1}]$	$L[10^{-4}, 10^{-1}]$	1.0	$U[0.5, 1.5]$

Table 1: Uncertainty assumptions

waiting for repair (tildes above the unit names) or being repaired (hats above the unit names). The CTMC of Figure 2 does not exhibit a product form solution; a closed form solution for the steady state distribution $\underline{\pi}(\lambda_A, \lambda_B, \phi) = [\pi_{(A,B)}, \pi_{(\widehat{A},B)}, \pi_{(A,\widehat{B})}, \pi_{(\widetilde{A},\widehat{B})}, \pi_{(\widehat{A},\widetilde{B})}]$ can still be given, as the matrix $\mathbf{Q}$ for this system is very small. We obtain that

$$\underline{\pi}(\lambda_A, \lambda_B, \phi) = C^* \cdot [\phi^2, \phi\lambda_A, \phi\lambda_B, \lambda_A\lambda_B, \lambda_A\lambda_B], \tag{26}$$

with $C^* = (2\lambda_A\lambda_B + \phi(\lambda_A + \lambda_B + \phi))^{-1}$.

8.3 The numerical assumptions

We analyse how uncertainty present in the parameters is propagated through the model to the performability of the system. Model parameters λ_A, λ_B and ϕ become random variables Λ_A, Λ_B and Φ. We do the analyses for five scenarios. For all these five cases, the rewards are chosen as follows: $r(A, B) = 1.0$, $r(\widehat{A}, B) = 0.75$, $r(A, \widehat{B}) = 0.25$, $r(\widetilde{A}, \widehat{B}) = r(\widehat{A}, \widetilde{B}) = 0.0$. Thus, it is assumed that processing unit A has a smaller processing capacity than processing unit B.

Further, in the scenarios $S1$–$S3$, it is assumed that processing unit A is less failure prone than processing unit B, and in the two remaining scenarios $S4$ and $S5$ that they are equally failure prone. The processing unit failure rates are based on expert knowledge. The "guesses" for the failure rates may therefore be correlated. We construct the dependence by means of a latent variable $\mathcal{L}$ to which both failure rates are coupled. Scenarios $S1$–$S3$ differ only with respect to this degree of coupling. The same holds for $S4$ and $S5$. The repair rate Φ is assumed to be independent of the failure rates.

Table 1 shows the numerical choices we consider for the distributions and rank correlations. The rank correlations ρ_r between Λ_A and Λ_B are obtained by coupling them each with the square root of this value to a latent variable $\mathcal{L}$. From Table 1 we calculate the values in Table 2, where $\sigma[X]$ denotes the standard deviation of random variable X and where $v[X]$ is the coefficient of variation of X:

$$v[X] = \sigma[X]/E[X]. \tag{27}$$

Λ_A	$E[\Lambda_A]$	$\sigma[\Lambda_A]$	$v[\Lambda_A]$
$S1$–$S3$	$2.150 \cdot 10^{-4}$	$2.497 \cdot 10^{-4}$	1.162
$S4, S5$	$1.446 \cdot 10^{-2}$	$2.269 \cdot 10^{-2}$	1.569
Λ_B	$E[\Lambda_B]$	$\sigma[\Lambda_B]$	$v[\Lambda_B]$
$S1$–$S3$	$2.150 \cdot 10^{-3}$	$2.497 \cdot 10^{-3}$	1.162
$S4, S5$	$1.446 \cdot 10^{-2}$	$2.269 \cdot 10^{-2}$	1.569
Φ	$E[\Phi]$	$\sigma[\Phi]$	$v[\Phi]$
$S1$–$S5$	1.000	.2887	.2887

Table 2: Summary of marginal parameter uncertainties

	Y	$\partial Y/\partial\Lambda_A$	$\partial Y/\partial\Lambda_B$	$\partial Y/\partial\Phi$
$S1$–$S3$	0.9983	-0.2520	-0.7470	$1.660 \cdot 10^{-3}$
$S4, S5$	0.9856	-0.2565	-0.7423	$1.444 \cdot 10^{-3}$

Table 3: The performability Y for the five average scenarios

We adopt the coefficient of variation $v[X]$ as a measure for the variability of X ($E[X] \neq 0$).

8.4 The sensitivity analysis

To perform a sensitivity analysis for the 5-state model, we take the values for the means and standard deviations from Table 2. Furthermore, we need the covariance values between the failure rates Λ_A and Λ_B. It is quite tiresome to compute these for the scenarios $S2, S3$ and $S5$. We avoid doing so by using in the sample covariance values obtained from the Monte Carlo simulations discussed in the next subsection. They equal $2.288 \cdot 10^{-7}$, $4.988 \cdot 10^{-7}$ and $5.147 \cdot 10^{-4}$ for the scenarios $S2$, $S3$ and $S5$ respectively.

Combining the steady state solution (26) with the rewards r, the expression for the performability $Y(\Lambda_A, \Lambda_B, \Phi)$ of the model equals

$$Y(\Lambda_A, \Lambda_B, \Phi) = \frac{\Phi(0.75\Lambda_A + 0.25\Lambda_B + \Phi)}{2\Lambda_A\Lambda_B + \Phi(\Lambda_A + \Lambda_B + \Phi)}. \tag{28}$$

Substitution of the values corresponding to the five mean scenarios of Table 2 in (28) and in the expressions for its partial derivatives, yields the values in Table 3.

By combining the values reported in Tables 2 and 3 according to (18) we compute an approximation for $\sigma[Y]$ around the performability of the average scenarios Y. The results are shown in Table 4.

Note that, as we investigate highly dependable systems, the performability Y always starts with 0.9 or 0.99; i.e. Y does not start with significant digits. If we simply calculate the coefficient of variation $v[Y]$ then we do not get a number that repre-

	$Y(E[\Lambda])$	$\sigma[Y]$	$v[Y]$	$v[1-Y]$
$S1$	0.9983	$1.927 \cdot 10^{-3}$	$1.930 \cdot 10^{-3}$	1.159
$S2$	0.9983	$1.938 \cdot 10^{-3}$	$1.941 \cdot 10^{-3}$	1.165
$S3$	0.9983	$1.951 \cdot 10^{-3}$	$1.954 \cdot 10^{-3}$	1.173
$S4$	0.9856	$1.830 \cdot 10^{-2}$	$1.857 \cdot 10^{-2}$	1.267
$S5$	0.9856	$2.304 \cdot 10^{-2}$	$2.338 \cdot 10^{-2}$	1.596

Table 4: Uncertainty propagation by sensitivity analysis

sents the amount of variability present in the digits of Y we are interested in. Therefore, for the performability Y, we use as measure of variability $v[1-Y] = \sigma[1-Y]/(E[1-Y]) = \sigma[Y]/(1-E[Y])$ (we use here that $\text{var}[a+bX] = b^2\text{var}[X]$).

We can interpret the resulting uncertainty in Y induced by the uncertainties and dependencies in the model parameters. Table 4 shows that the performability in the scenarios $S1$–$S3$ is better than the performability for $S4$ and $S5$. This is of course due to the fact that one of the machines has a much smaller failure rate in the former scenarios. Furthermore, we can conclude that the standard deviation for $S1$–$S3$ only depends to a very small extend on the dependence between the two failure rates. This is due to the fact that one of the failure rates dominates the entire performability of the system. That more dependence between the failure rates implies a higher standard deviation in the performability is better illustrated by scenarios $S4$ and $S5$ where the failure rates are completely independent respectively dependent, and of the same order of magnitude.

We do not only want to know the "variability" in the model output, but we also want to compare it with the variability introduced in the model parameters. We want to get an indication whether the model increases or decreases uncertainty present in the model parameters by the propagation to the system performability. The results for $v[1-Y]$ indicate that the variability of the performability Y, restricted to the digits that are of interest, is of the same order of magnitude as the variability of the failure rates Λ_A and Λ_B.

8.5 The uncertainty analysis

For the five scenarios $S1$–$S5$ we generated $m = 10000$ samples each, using the Unicorn package developed at the University of Delft [6]. These samples where fed to a program that evaluated the sketched CTMC, yielding 10000 samples $y_1, \cdots, y_m$ of the steady state performability measure Y. Results on the main sample moments Y are presented in Table 5. The sample mean of stochastic variable X is denoted by $\hat{\mu}[X]$, its

	$\hat{\mu}[Y]$	$\hat{\sigma}[Y]$	$\hat{v}[Y]$	$\hat{v}[1-Y]$
S1	.9982	$2.197 \cdot 10^{-3}$	$2.201 \cdot 10^{-3}$	1.207
S2	.9981	$2.304 \cdot 10^{-3}$	$2.308 \cdot 10^{-3}$	1.245
S3	.9981	$2.310 \cdot 10^{-3}$	$2.314 \cdot 10^{-3}$	1.248
S4	.9848	$1.964 \cdot 10^{-2}$	$1.994 \cdot 10^{-2}$	1.290
S5	.9839	$2.653 \cdot 10^{-2}$	$2.696 \cdot 10^{-2}$	1.652

Table 5: Uncertainty propagation by uncertainty analysis: moments

	S1	S2	S3	S4	S5
5% q.	.9935	.9932	.9932	.9431	.9261
50% q.	.9992	.9991	.9992	.9928	.9966
95% q.	.9998	.9999	.9999	.9997	.9999

Table 6: Uncertainty propagation by uncertainty analysis: quantiles

sample standard deviation by $\hat{\sigma}[X]$ and the ratio of these two numbers is denoted by $\hat{v}[X]$. The remarks that have been made for the sensitivity analysis are also applicable here.

Results on the quantiles of Y are presented in Table 6. It can be observed that the 50% quantile is much closer to the 95% quantile than to the 5% quantile. This indicates that the distribution of Y is skewed to the left. This can also be seen from the fact that the mean performability $\hat{\mu}[Y]$ is much smaller than the 50% quantile. This skewness can also be observed in Figures 3 and 4 which show $\Pr\{Y \leq y\}$ against y.

The fact that the failure rate dependence does not affect Y very much in the scenarios $S1$–$S3$ can also be observed in Figure 3 (the curves are overlapping). The effect of the dependence is somewhat stronger in scenarios $S4$ and $S5$ (see Figure 4). This is also confirmed by the rank-correlations in Table 7. The rank-correlation between Λ_B and Y is close to -1 in scenarios $S1$–$S3$ because the system performability is dominated by the well-operation of component B. The rank-correlation between the repair rate Φ and Y is always around 0.22. When the dependence between Λ_A and Λ_B increases, the rank-correlation between Λ_A and Y also increases. This does not indicate that Λ_A influences Y more, but merely that Λ_A is coupled more tightly to Λ_B.

In scenario $S4$, the difference between the rank-correlation between Λ_A and Y and Λ_B and Y can be explained by the difference between the rewards that are obtained when either component A or B is operational. Going from $S4$ to $S5$, which means increasing the dependence between the failure rates, simply increases the rank-correlations. This means that in case of dependent failure rates, Y strongly depends on the failure rates and much less on the repair rate.

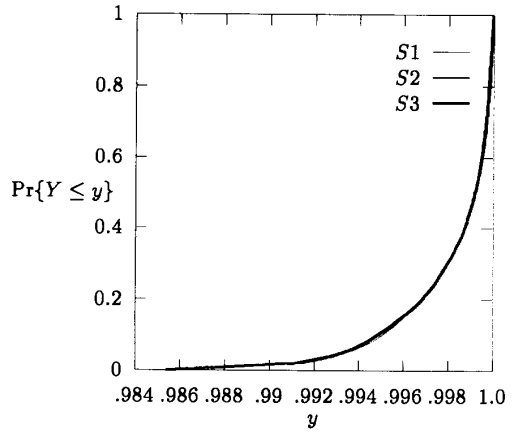

Figure 3: $\Pr\{Y \leq y\}$ against y as derived by $S1$, $S2$, and $S3$

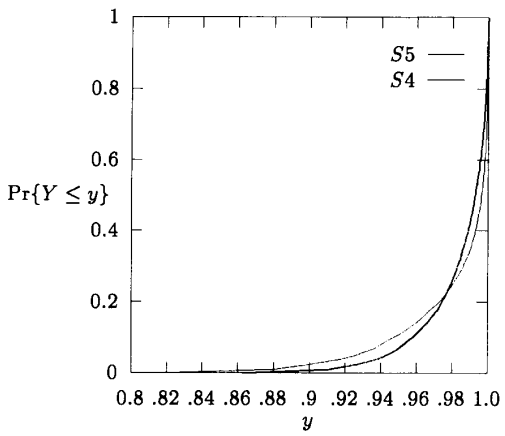

Figure 4: $\Pr\{Y \leq y\}$ against y as derived by $S4$ and $S5$

	$\rho_r(Y, \Lambda_A)$	$\rho_r(Y, \Lambda_B)$	$\rho_r(Y, \Phi)$
$S1$	-0.075	-0.963	0.266
$S2$	-0.538	-0.971	0.222
$S3$	-0.892	-0.975	0.218
$S4$	-0.425	-0.794	0.174
$S5$	-0.989	-0.989	0.157

Table 7: Rank-correlations between the parameters Λ_A, Λ_B, Φ and the performability Y

8.6 The evaluation

The derived results of the sensitivity analysis and of the uncertainty analysis agree fairly well. Important to observe is that the results from the sensitivity analysis compared to those obtained with the uncertainty analysis, for all five scenarios $S1$–$S5$, *over*-estimate the mean performability, and *under*-estimate its standard deviation. We think this is due to the fact that the sensitivity analysis approach is based on the mean and variance of the parameter distributions and the first derivative of Y with respect to the model parameters only. It therefore implicitly assumes a symmetric probability distribution for Y, which is not correct. The three quantiles that can be derived by an uncertainty analysis describe the results better; they do reflect the skewness (a skew distribution is not described sufficiently by its first two moments).

9 Summary and conclusions

In this paper we have investigated how model parameter uncertainty is propagated to performability measures derived from the model. We have followed two approaches to do this: an approach based on parametric sensitivity analysis and an approach based on Monte Carlo uncertainty analysis.

In the application we considered, the sensitivity analysis consequently over-estimated the average performability and under-estimated the variance in the performability due to the parameter uncertainties. This behaviour can also be observed in other, much larger, case studies we did [11].

In this paper we have presented the performability models as Markov reward models. Preferably however, we specify the models at a higher level of abstraction and let the "translation" to the lower level, the solution of the lower level model as well as the enhancement of the results to the higher level be done automatically. In [11] (of which this paper is a part) we have used the dynamic queueing network concept [9], implemented on top of stochastic reward nets [3], to describe the performability models. As tool we have used SPNP [2] since it allows for the easy description of parameterized models and it allows for the automatic derivation of first order derivatives of the performability measure of interest.

In the near future, we will extend our work in a number of directions. We will apply both approaches to a wider set of scenarios and to a wider class of

models. For the sensitivity approach, we will generalize the one-dimensional Taylor series expansion to the multi-dimensional case [10]. We will also work on an approach to incorporated higher moments of the parameter uncertainty distribution in the sensitivity analysis.

References

[1] J.T. Blake, A.L. Reibman, K.S. Trivedi, "Sensitivity Analysis of Reliability and Performance Measures for Multiprocessor Systems", *ACM Performance Evaluation Review* **16**(1), pp.177–186, 1988.

[2] G. Ciardo, J. Muppala, K.S. Trivedi, "SPNP: Stochastic Perti Net Package", *Proceedings PNPM89*, IEEE Computer Society Press, pp.142–151, 1989.

[3] G. Ciardo, J. Muppula, K.S. Trivedi, "On the Solution of GSPN Reward Models", *Performance Evaluation* **12**, pp.237–253, 1991.

[4] R.M. Cooke and R. Waij, "Monte Carlo Sampling for Generalized Knowledge Dependence with Application to Human Reliability", *Risk Analysis* **6**(3), pp.335-343, 1986.

[5] N.M. van Dijk, "Transient Error Bound Analysis for Continuous Time Markov Reward Structures", *Performance Evaluation* **13**, pp.147–158, 1991.

[6] J.R. van Dorp, *Dependence Modeling for Uncertainty Analysis*, Delft University of Technology, december 1991.

[7] V. Grassi, L. Donatiello, "Sensitivity Analysis of Performability", *Performance Evaluation* **14**, pp.227–237, 1992.

[8] Y.-C. Ho, X.-R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems*, Kluwer Academic Publishers, 1991.

[9] B.R. Haverkort, I.G. Niemegeers, "Performability Modelling Tools: A Survey", *Memoranda Informatica* 91-82, University of Twente, 1991.

[10] B.R. Haverkort, A.M.H. Meeuwissen, "Multi-Dimensional Sensitivity Analysis of Performability", in progress, 1992.

[11] B.R. Haverkort, A.M.H. Meeuwissen, "Sensitivity and Uncertainty Analysis in Performability Modelling", *Memoranda Informatica* 92-18, 1992.

[12] P. Heidelberger, A. Goyal, "Sensitivity Analysis of Continuous Time Markov Chains using Uniformization", in *Computer Performance and Reliability*, Editors: G. Iazeolla, P.-J. Courtois, O.J. Boxma, North-Holland, pp.93–104, 1988.

[13] R.L. Iman, W.J. Conover, "A Distribution-Free Approach to Inducing Rank Correlation among Input Variables", *Communications in Statistics-Simulation and Computation* **11**(3), pp.311–334, 1982.

[14] R.L. Iman, J.C. Helton, "An Investigation of Uncertainty and Sensitivity Analysis Techniques for Computer Models", *Risk Analysis* **8**(1), pp.71–90, 1988.

[15] M.E. Johnson, *Multi-Variate Statistical Simulation*, Wiley, 1987.

[16] H.S. Konijn, *Statistical Theory of Sample Survey Design and Analysis*, North-Holland, 1973.

[17] B. Krzykacz, E. Hofer, "The Generation of Experimental Designs for Uncertainty and Sensitivity Analysis of Model Predictions with Emphasis on Dependences between Uncertain Parameters", in: *Reliability of Radioactive Transfer Models*, Editor: G. Desmet, North-Holland, 1988.

[18] S. Kullback, *Information Theory and Statistics*, Wiley, 1959.

[19] J.C. Laprie, "Dependable Computing and Fault-Tolerance: Concepts and Terminolgy", *Proceedings FTCS 15*, IEEE Computer society Press, pp.2–7, 1985.

[20] A.M.H. Meeuwissen, "A Latent Variable Model for Knowledge Dependence", Delft University of Technology, in progress, 1992.

[21] A.M.H. Meeuwissen, T. Bedford, "Some Notes on Joint Distributions with Maximal Entropy given Rank Correlation", Delft University of Technology, in progress, 1992.

[22] M.G. Morgan, M. Henrion, *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, Cambridge University Press, 1990.

[23] N.M. van Dijk, B.R. Haverkort, I.G. Niemegeers (editors), "Guest Editorial: Performability Modelling of Computer and Communication Systems", *Performance Evaluation* **14**, pp.135–138, 1992.

[24] P.D. Welch, "The Statistical Analysis of Simulation Results", in *Computer Performance Modelling Handbook*, Editor: S.S. Lavenberg, Academic Press, New York, pp.267–329, 1983.