



## Second-order Monte Carlo uncertainty/variability analysis using correlated model parameters: application to salmonid embryo survival risk assessment

Fu-Chun Wu<sup>a,b,\*</sup>, Yin-Phan Tsang<sup>a</sup>

<sup>a</sup> Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei 106, Taiwan, ROC

<sup>b</sup> Hydrotech Research Institute, National Taiwan University, Taipei 106, Taiwan, ROC

Received 12 September 2003; received in revised form 17 February 2004; accepted 26 February 2004

### Abstract

This paper presents a Monte Carlo uncertainty/variability analysis applied to assessment of salmonid embryo survival affected by the deposition of fine sand in the spawning gravel riverbed. For the system being modeled, the uncertainty originates from the lack of perfect knowledge about the spawning habitat, while the variability arises from the natural diversity in the amount of sand deposit. A second-order Monte Carlo simulation (MCS) procedure consisting of multiple realizations of model parameters and full-range iterations of input variable is used to separately propagate the knowledge uncertainty and natural variability of the system. Given that the three parameters of the present model are highly correlated (correlation coefficients >0.8), we use four multivariate methods (Iman–Conover method, standard normal transformation method, normal copula method, and maximum-entropy copula method) to generate correlated model parameters. The scatter plots of the correlated parameters generated with various methods demonstrate different correlation patterns. An important finding of the present uncertainty/variability analyses and risk assessments is that the outcomes resulting from the four correlated scenarios are very similar, regardless of the diverse correlation patterns. The output of the second-order MCS is a collection of cause–effect relations (spaghetti plot) that can facilitate the analyses of survival-rate uncertainty/variability. The favorable habitat conditions are also defined on the basis of the spaghetti plot. It is revealed that significant values of information are associated with the parameter correlations, especially when higher levels of embryo survival are targeted. The results further indicate that the outcomes of the independent (less realistic) scenario substantially differ from those of the correlated ones, indicating that the effect of parameter correlations should be incorporated into the MCS whenever possible. This work provides a useful paradigm for the systematic uncertainty/variability analyses in the context of ecological risk assessments.

© 2004 Elsevier B.V. All rights reserved.

**Keywords:** Monte Carlo simulation; Uncertainty/variability analysis; Ecological risk assessment; Model parameter; Correlation; Salmonid embryo survival; Spawning habitat

### 1. Introduction

The total uncertainty (unpredictability) of a system is the combination of two components, namely, uncertainty and variability (Hattis and Burmaster, 1994; Rai

\* Corresponding author. Tel.: +886-2-2369-2466;  
fax: +886-2-2369-2466.  
E-mail address: [fcwu@hy.ntu.edu.tw](mailto:fcwu@hy.ntu.edu.tw) (F.-C. Wu).

et al., 1996; Vose, 2000). Uncertainty is the analyst's lack of knowledge (or level of ignorance) about the system, which may be reduced by further measurement or study. Variability represents the random (or stochastic) heterogeneity in a population (people or events) that is irreducible by additional measurements or studies but may be reduced by changing the system (Anderson and Hattis, 1999; Vose, 2000). Both uncertainty and variability can be described by probability density functions (PDFs), one might therefore mix them in the Monte Carlo risk assessment (in which some distributions reflect the uncertainty about model parameters, while the others reflect the inherent stochastic nature of the system) and claim that the output would take account of all uncertainty and variability. However, the resultant single distribution is technically difficult to interpret and information is lost regarding the individual contributions of uncertainty and variability to the total uncertainty. Moreover, a much larger problem than mixing uncertainty and variability can occur when a variability distribution is used as if it were an uncertainty distribution, in this case a variable could be mistakenly treated as an uncertain parameter such that the output that should be a single figure is replaced by a meaningless distribution (Vose, 2000). Therefore, it has been increasingly emphasized the need to separate uncertainty and variability in risk assessments (Hoffman and Hammonds, 1994; Frey and Rhodes, 1996; Rai and Krewski, 1998; Nauta, 2000; Mitchell and Csillag, 2001). The separation of uncertainty and variability allows us to identify the steps that can be taken to reduce the total uncertainty, to gauge the value of information or of potential system changes, and to avoid the serious errors induced by mixing variables with parameters.

In recent years, the second-order Monte Carlo simulation (MCS), also known as two-phase or two-dimensional MCS, has been extensively used to separately propagate uncertainty and variability in risk assessments (e.g., Bogen and Spear, 1987; Hoffman and Hammonds, 1994; Hession et al., 1996; Burmaster and Wilson, 1996; Nauta, 2000; Vose, 2000; Sanga et al., 2001; Moschandreas and Karuchit, 2002; Pouillot et al., 2003). The second-order MCS involves a double looping (or nesting) procedure consisting of multiple realizations of model parameters and iterations of input variables. The outcome

of the second-order MCS is a collection of cumulative distribution functions (CDFs) or cause–effect relations that simultaneously display the uncertainty and variability in the results. In this study, we apply the second-order MCS in the context of salmonid embryo survival risk assessment. Based on the results of such analyses, the implications to the management of spawning habitat are also explored.

Parameter distributions in a model often have to be correlated in order to ensure that only meaningful scenarios are simulated (Vose, 2000; Todd and Ng, 2001). It has been recognized that when the parameters of a model are correlated, incorporation of such correlations in the MCS is important, especially when the correlation coefficients exceed 0.7 (Smith et al., 1992; Bukowski et al., 1995; Haas, 1999; Pohlmann et al., 2002). The correlation (or dependence) between the random numbers is generally indicated by such measures as the product-moment correlation coefficient (Pearson's  $r$ ) or rank–order correlation coefficient (Spearman's  $\rho$  or Kendall's  $\tau$ ) (see, e.g., Nelsen, 1999 and Vose, 2000). Parameter correlations are incorporated into the MCS by generating many sets of correlated model parameters and using different set in each realization. In this study, we employ four multivariate methods to generate correlated model parameters. The parameter vectors produced with these methods are compared. Further, the effect of parameter correlations and the associated value of information are evaluated. These results provide valuable insights into the problem and useful guidelines for habitat improvement.

Monte Carlo methods have been widely used in the context of ecological modeling (Neave et al., 1997; Harmon and Challenor, 1997; Annan, 2001; Qian et al., 2003; Waller et al., 2003). However, Mitchell and Csillag (2001) pointed out that ecosystem models that include both variability of driving variables and uncertainty in their predictions are rare. The second-order MCS is employed herein to address this issue. Moreover, various model approaches to survival risk assessment have been presented, such as the stochastic model (Todd and Ng, 2001) and vitality-based model (Salinger et al., 2003). The present survival model is based on deterministic cause–effect relations validated by field data, thus eliminates the process-based stochasticity.

## 2. Salmonid embryo survival model

Field studies have indicated that salmonids (salmon and trout) explore suitable locations in the gravel riverbed for spawning, where the female excavates a pit and releases fertilized eggs into the bottom. After spawning, the female resumes digging upstream to bury the area of egg deposition (so called “egg pocket”). The embryos are therein protected against bedload motion and scour during high flows. The typical morphology of the resultant nest (redd), with a pit upstream and a tailspill downstream, is shown in Fig. 1. The hydrodynamics induced by the bed topography promotes a downwelling flow into redd. The pressure gradient exerted between the upstream and downstream faces of the tailspill forces streamwater to flow through the gravels. This flow benefits the exchange of dissolved oxygen (DO) and removal of metabolic wastes (MW) such that a suitable intragravel environment for embryo development is maintained. Among the factors adversely affecting embryo survival, accumulation of fine sand in the gravel substrate is one of the most detrimental. Fishery researchers generally agree that intrusion of sand into spawning gravels substantially reduces substrate permeability and intragravel water velocities, thereby restricting supply of DO and removal of MW. Modeling of embryo survival is based on the above-stated mechanisms (for details see Wu, 2000), which were validated by field data (Tappel and Bjornn, 1983). A brief overview of this model is as follows.

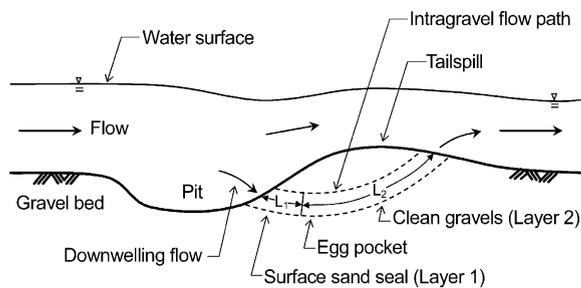


Fig. 1. Typical morphology of salmonid spawning nest (redd), with a pit upstream and a tailspill downstream. The bed topography promotes a downwelling flow into redd.  $L_1$  is the length of flow path through surface sand seal (layer 1),  $L_2$  is the length of flow path through surrounding gravels (layer 2) (modified from Wu, 2000).

### 2.1. Model components

#### 2.1.1. Relation between sand deposit and substrate permeability

A nonlinear relation between the reduced substrate permeability and sand content is given by the following expression derived from the hydraulic radius theory (Wu, 2000):

$$\frac{K}{K_0} = 4.54 \times \frac{(0.42 - 1.54\sigma)^3}{(0.58 + 1.54\sigma)^2} + 3.66 \left( \frac{d_s}{D_g} \right)^2 \sigma \quad (1)$$

where  $K_0$  is the permeability of the clean gravels;  $K$  is the reduced permeability of the polluted gravels;  $\sigma$  is the specific sand deposit, defined as the volume ratio of sand deposit to bulk gravels;  $d_s/D_g$  is the ratio of sand to gravel grain sizes. For a clean gravel bed ( $\sigma = 0$ ), the second term on the RHS of Eq. (1) is ineffective, whereas the first term vanishes as the intra-gravel pores are saturated with sand (corresponding to  $\sigma = 0.273$ ).

#### 2.1.2. Relation between substrate permeability and seepage velocity

Streamwater enters the redd gravels in the high-pressure region (upstream face of tailspill) and leaves the gravel bed in the low-pressure region (downstream face of tailspill). The typical flow path is illustrated in Fig. 1. The substrate gravels tend to filter out the sand particles carried in the water, thus forming a sand seal in the surface layer of the gravel bed that inhibits further sand infiltration. The permeability of the sand seal is relatively low compared to the surrounding gravels. On the basis of Darcy’s law, the apparent seepage velocity through the two-layer redd gravels can be determined by:

$$V = \frac{(h/L_1)K_2}{L_2/L_1 + K_2/K_1} \quad (2)$$

where  $L_1$  is the length of flow path through layer 1 (sand seal);  $L_2$  is the length of flow path through layer 2 (surrounding gravels);  $K_1$  and  $K_2$  are the permeability of layers 1 and 2, respectively;  $h$  is the total pressure-head drop between the upstream and downstream faces of tailspill. The ratio  $K_1/K_2$  used in Eq. (2) is simply the  $K/K_0$  value obtained from Eq. (1), for which  $\sigma$  is taken to be the specific deposit in layer 1.

### 2.1.3. Relation between seepage velocity and embryo survival

The apparent seepage velocity can serve as an indicator for embryo survival assessment. The experimental data of Cooper (1965) were used to develop an empirical relation between seepage velocity and embryo survival rate ( $R^2 = 0.99$ ,  $n = 8$ ):

$$S = -17.6(\log V)^2 - 39.6(\log V) + 68.7 \quad (3)$$

where  $S$  is percentage survival;  $V$  is in cm/s.

## 2.2. Model parameters

### 2.2.1. PDFs of model parameters

The embryo survival model described above involves three physical parameters, namely, grain-size ratio  $d_s/D_g$ , pressure-head ratio  $h/L_1$ , and flow-path ratio  $L_2/L_1$ . To determine these values, Chinook salmon (*Oncorhynchus tshawytscha*) is selected as the target species because more data are available for this species. The ranges of these parameters and their most probable values, listed in Table 1, were obtained through a comprehensive survey of literature (Wu, 2000). However, the currently available data are insufficient to determine the PDFs of these parameters. Normal and log normal distributions are among the most widely adopted PDFs for modeling uncertain parameters, but they are unbounded at two and one side, respectively, thus are inappropriate for the bounded parameters. Nevertheless, since we have the possible ranges and the most probable values of parameters, it seems reasonable to use the bounded distributions, such as triangular distributions, as a first approxima-

Table 1

Ranges and most probable values of habitat characteristics and model parameters (modified from Wu, 2000)

Item	Range	Most probable value
Habitat characteristic	Observed data range	Most frequently observed
$D_g/d_s$	7–30	14
$h$	1–10 cm	4 cm
$L_1$	5–10 cm	8 cm
$L_2$	100–350 cm	280 cm
Model parameter	Inferred parameter range	Inferred
$d_s/D_g$	0.03–0.14	0.07 (=1/14)
$h/L_1$	0.1–2	0.5 (=4 cm/8 cm)
$L_2/L_1$	10–70	35 (=280 cm/8 cm)

Target species: Chinook salmon (*Oncorhynchus tshawytscha*).

tion to the PDFs of model parameters. The minimum and maximum observed values are used to infer the lower and upper bounds of the triangular PDFs, and the most probable values are taken to be the modes of the PDFs (Fig. 2). Although the observed ranges of parameter values are used herein, there is no support for setting such limits on the ranges of uncertainty, thus one could allow slightly wider ranges if desired.

### 2.2.2. Correlations among model parameters

The experimental data of sand infiltration into a gravel bed (Wu and Shen, 1999) are used to evaluate the correlations among the parameters. A total of 13 data sets are analyzed, resulting in a correlation matrix as given in Table 2. Because the calculated product-moment and rank-order correlation coefficients are very close, the values listed in Table 2 can

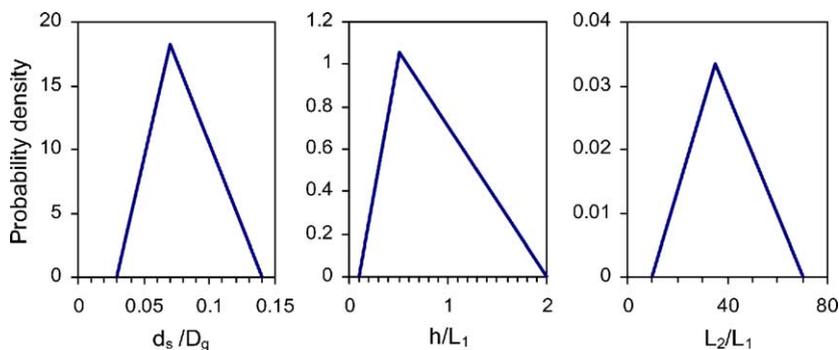


Fig. 2. Probability density functions (PDFs) of three model parameters. The minimum and maximum observed values are used to infer the lower and upper bounds, and the most probable values are taken to be the modes of triangular PDFs.

Table 2  
Correlation matrix of model parameters

	$d_s/D_g$	$h/L_1$	$L_2/L_1$
$d_s/D_g$	1	0.8	0.9
$h/L_1$		1	0.8
$L_2/L_1$			1

be used in either context. It turns out that the correlations among the parameters are all positive and greater than 0.8, with the most significant correlation existing between  $d_s/D_g$  and  $L_2/L_1$ . These values also suggest that we need to incorporate parameter correlations in the present study.

### 3. Generation of correlated random numbers

There are numerous ways of obtaining correlated random numbers (Burmester and Anderson, 1994). Herein, we employ four methods appropriate for multivariate correlations to generate a large number of parameter sets using the marginal PDFs and correlations specified in the preceding sections. These methods include a distribution-free approach that involves the orthogonal transformation of a van der Waerden random score matrix (i.e. Iman–Conover or IC method); an inverse scheme based on the orthogonal transformation in the standard normal (SN) space (i.e. standard normal transformation or SNT method); and two multivariate copula methods respectively employing the continuous normal and piecewise-uniform copula densities (i.e. normal copula and maximum-entropy copula or NC and MEC methods). The theoretical background and procedure of these methods are included in Appendix A.

For illustration, Fig. 3 presents the scatter plots of  $d_s/D_g$  versus  $L_2/L_1$  generated with various methods (with rank or moment correlation = 0.9), where different sample sizes are the result of convergence, goodness-of-fit, and correlation tests. The convergence test is to check if the means and standard deviations of the generated random numbers converge to the prescribed marginal values. If the test criterion is not met we generate 100 more observations, primarily due to the concern that the generated random numbers do not sufficiently span the parameter space. As such, among the four methods, only IC method can generate a fixed

sample size of parameter vectors meeting the specified criteria. The independent scatter plot and the experimental data (Wu and Shen, 1999) are also shown for comparison (Fig. 3), where the independent random numbers were generated using Latin hypercube sampling (LHS) technique. In those correlated scatter plots, an obvious positive-correlation scatter pattern is demonstrated, although their shapes are different. The IC, SNT, and NC scatter plots are elliptical in shape, whereas the MEC plot is piecewise “pinched” in four blocks. The elliptical-shaped scatter plot of IC method is the result of using van der Waerden scores in the generation of correlated random numbers, which was the original intention of Iman and Conover (1982) to use such scores for producing “natural-looking” correlations. The NC scatter plot appears to be in good agreement with the IC plot (Haas, 1999). However, due to the acceptance/rejection procedure employed, the NC method needed more iterations and computation time (in the order of several hours) to meet the specified criteria for the three tests. The sample size of SNT method was the greatest, making its scatter plot spindle in shape. Nonetheless, owing to the straightforward inverse transformation scheme employed, the computation time taken by the SNT method was in the order of one minute.

The piecewise pinched scatter plot of MEC method is unique among others, which is a combined result of partitioning each uniform marginal on the unit hypercube into four equal-size pieces for constructing the piecewise-uniform copula densities and the high degree of correlation adopted. For example, the random numbers in the second interval of  $d_s/D_g$  are associated with the majority of random numbers in the second interval of  $L_2/L_1$ , with only rare in the first and third intervals but none in the fourth. Note that the unequal-size intervals on both marginals of the MEC scatter plot are attributed to the transformation between uniform and triangular distributions. Although the result of MEC method does not look very realistic, it does provide proper correlation, convergence, and goodness of fit. While these matched moments are not sufficient enough to encourage use of this approach, as will be shown later on, all these four methods yield convergent results in the uncertainty/variability analyses and risk assessments.

The results shown in Fig. 3 indicate that the correlation pattern between the random variates can be

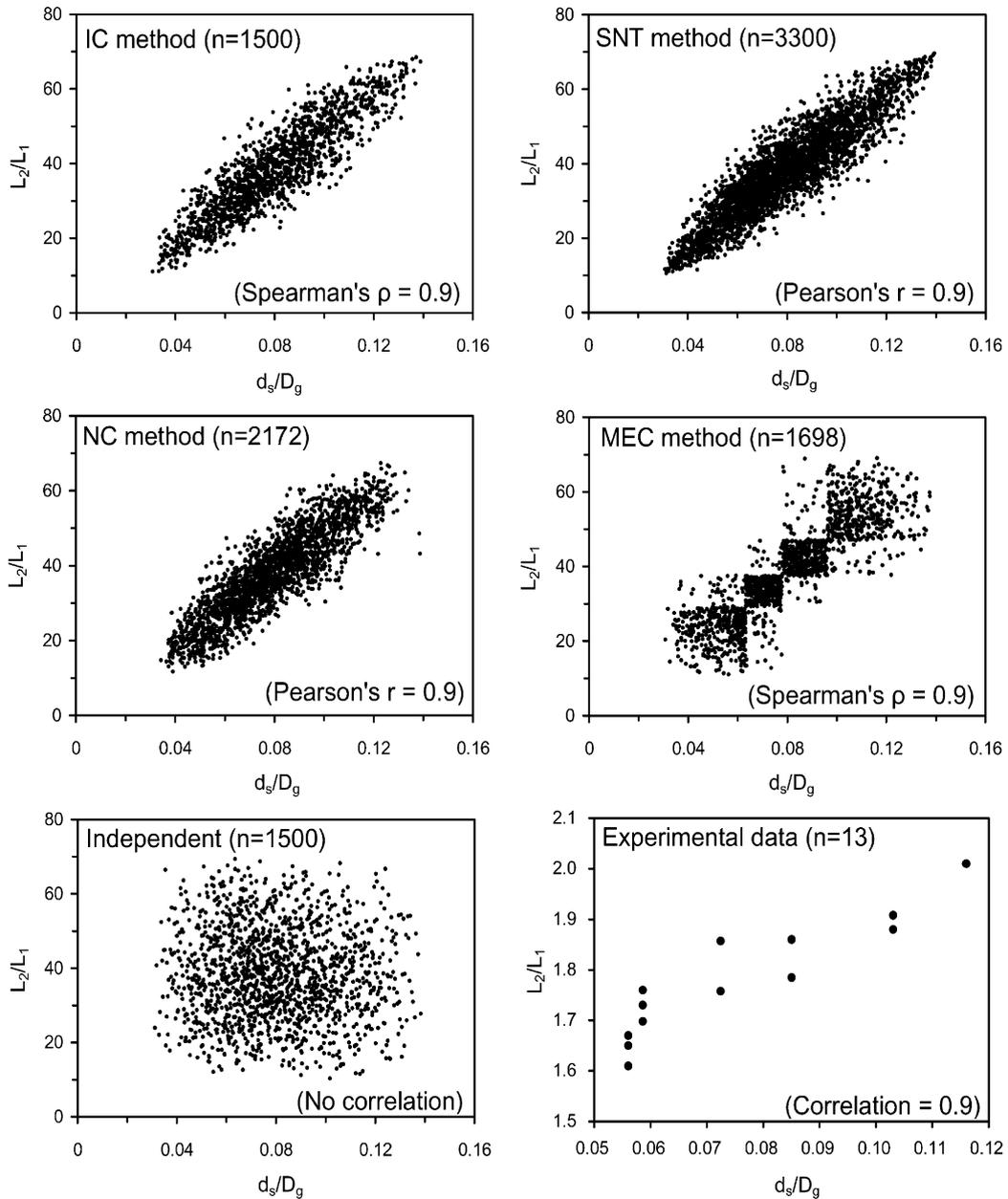


Fig. 3. Scatter plots showing the correlation between  $d_s/D_g$  and  $L_2/L_1$ . The random parameters are generated with four correlation methods (with rank or moment correlation = 0.9) and independent Latin hypercube sampling. Experimental data (Wu and Shen, 1999) are also shown.

rather different even specified with the same marginals and correlations. The correlation pattern observed in real data may not match with that produced (Fig. 3). Previous researchers have pointed out that to specify

only the marginal distributions and correlation coefficient might not be sufficient when it is desired to generate random numbers of a certain correlation pattern (e.g., Haas, 1999; Vose, 2000). The correlation

patterns among the present model parameters are beyond our current knowledge. Nevertheless, the effects of using different correlation methods on the outputs of various convolutions and, of more direct interest, on the outcomes of Monte Carlo uncertainty/variability analyses and risk assessments are investigated in this study.

#### 4. Second-order Monte Carlo simulation

The salmonid embryo survival model presented in Section 2 and the correlated model parameters obtained in Section 3 are integrated into the framework of second-order MCS to propagate the knowledge uncertainty and system variability. The components of uncertainty and variability involved and the procedure of second-order MCS are described below.

For the salmonid embryo survival model, the parameter uncertainty originates from the lack of perfect knowledge about the physical conditions of the spawning habitat, which include the gravel/sand grain sizes, flow hydraulics, and redd dimensions. Such knowledge uncertainty may be reduced through further

investigations. As mentioned earlier, with the current knowledge these parameter uncertainties are approximated by the triangular PDFs. Accordingly, sufficient sets of correlated model parameters have been generated; each set represents a meaningful and possible combination of habitat conditions. Propagation of such uncertainty is implemented by using a different set of parameters in each realization (Fig. 4). On the other hand, the amount of sand accumulated in the bed surface represents a natural diversity of the system, which varies as a complex function of many temporal and spatial factors (Wu and Shen, 1999; Wu, 2000). Such diversity is irreducible through further studies, and is the cause of survival-rate variability under a given combination of habitat conditions. Propagation of such variability is implemented by running multiple iterations with the input  $\sigma$ -value incrementally varying from 0 (clean gravels) to 0.273 (saturated sand seal) to cover the full range of specific sand deposit (Fig. 4).

The double-looping procedure of the second-order MCS is demonstrated in Fig. 4, where the outer loop consists of  $n$  realizations of model parameters to propagate the knowledge uncertainty; the inner loop consists of  $m$  iterations of input variable to propagate

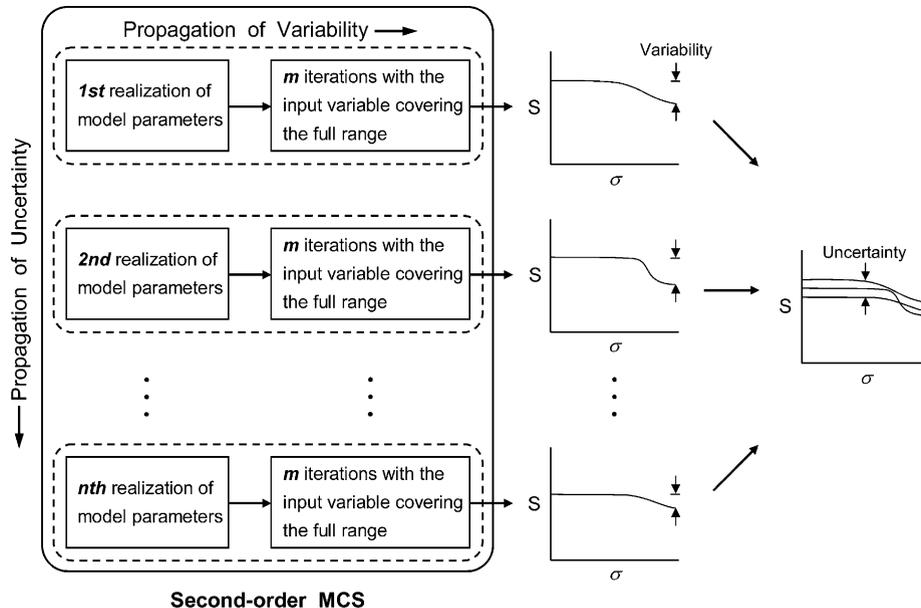


Fig. 4. Schematic diagram showing the procedure of second-order MCS and the outcomes. The inner loop consists of  $m$  iterations propagating the system variability; the outer loop consists of  $n$  realizations propagating the parameter uncertainty.

the system variability. The ultimate outcome of the second-order MCS is a collection of  $S$  versus  $\sigma$  relations simultaneously displaying the survival-rate uncertainty and variability. The results concerning the uncertainty and variability of embryo survival, the risk arising from knowledge uncertainty and system variability, and the value of information associated with parameter correlations are discussed in subsequent sections.

## 5. Uncertainty/variability analyses and risk assessments

The outcomes of the second-order MCS using the parameters generated with different correlation methods are shown in Fig. 5, where the “spaghetti-looking” graph demonstrates a number of possible  $S$  versus  $\sigma$  relations, each one of which corresponds to a different habitat scenario. On the basis of these results, the uncertainty/variability analyses and quantitative risk assessments can be conducted, and the effect of parameter correlations on the model outcomes can be evaluated. Currently a standard procedure for uncertainty/variability analyses is not yet available, although the second-order MCS is increasingly employed as a practical means to separately propagate uncertainty and variability. In this paper we systematically address these issues in an attempt to provide a paradigm for such analyses.

### 5.1. Comparison of the four methods for generating correlated random numbers

In this section, the uncertainty and variability of salmonid embryo survival are investigated, and the outcomes resulting from different correlated scenarios are compared. The four methods for generating correlated random numbers are evaluated on the basis of these results.

#### 5.1.1. Uncertainty of embryo survival rate

For each spaghetti plot in Fig. 5, the uncertainty of embryo survival rate at a given  $\sigma$ , denoted as  $\Delta S_u(\sigma)$ , is evaluated by  $S_{\max}(\sigma) - S_{\min}(\sigma)$ , where  $S_{\max}(\sigma)$  and  $S_{\min}(\sigma)$  are the maximum and minimum values of  $S$  corresponding to the given  $\sigma$ , respectively (as shown in Fig. 4). The survival-rate uncertainty  $\Delta S_u(\sigma)$  gen-

erally increases with  $\sigma$ . However, diverse results of  $\Delta S_u(\sigma)$  are associated with different correlation methods. For lower values of  $\sigma$  (roughly for  $\sigma < 0.15$ ), the SNT uncertainty is the smallest; the NC and MEC uncertainties are similar in magnitude and greater than other correlated ones. The survival-rate uncertainty associated with the independent parameters is consistently the largest among all results. These appear to indicate that, for those lower  $\sigma$ -values, incorporating parameter correlations into MCS reduces the uncertainty of the outcome due to limiting attention to only realistic combinations of parameter values. However, at higher values of  $\sigma$  (roughly for  $\sigma > 0.15$ ), no consistent trend can be found. The spaghetti plot of the independent parameters appears less orderly than the correlated ones, implying that some unrealistic or meaningless scenarios could be generated. Although the spaghetti plots of the correlated parameters appear more orderly, some extreme conditions might have arisen from the random sampling procedure. For example, in the NC spaghetti plot, the lowest  $S$  versus  $\sigma$  curve isolated from others is a good illustration of such situation. Herein, the concept of uncertainty interval (Sanga et al., 2001) is adopted to address this issue.

To eliminate the extreme effect, the 5th and 95th percentile values of  $S$ , along with the median  $S$ , are plotted against  $\sigma$  in Fig. 6a for the outcomes of different correlation methods. The difference between the 95th and 5th percentile values of  $S$  is the length of the 90% uncertainty interval (not to be confused with confidence interval), which is plotted against  $\sigma$  in Fig. 6b for different methods. It is revealed in Fig. 6a that the 5th, 95th percentile and median  $S$  versus  $\sigma$  curves of all correlation methods are in good agreement. The independent 95th percentile and median  $S$  versus  $\sigma$  curves coincide well with those correlated ones. However, the independent 5th percentile  $S$  versus  $\sigma$  curve is quite different from the correlated ones. The independent 5th percentile values of  $S$  are consistently smaller than the correlated ones for  $\sigma < 0.2$ , but then exceed the correlated ones for  $\sigma > 0.2$ . If the correlated scenarios are considered as more realistic, the implication of the results shown in Fig. 6b would be that by ignoring the parameter correlations the survival-rate uncertainty is overestimated for those less harmful conditions of sand deposit but is underestimated for those worst sand deposit conditions.

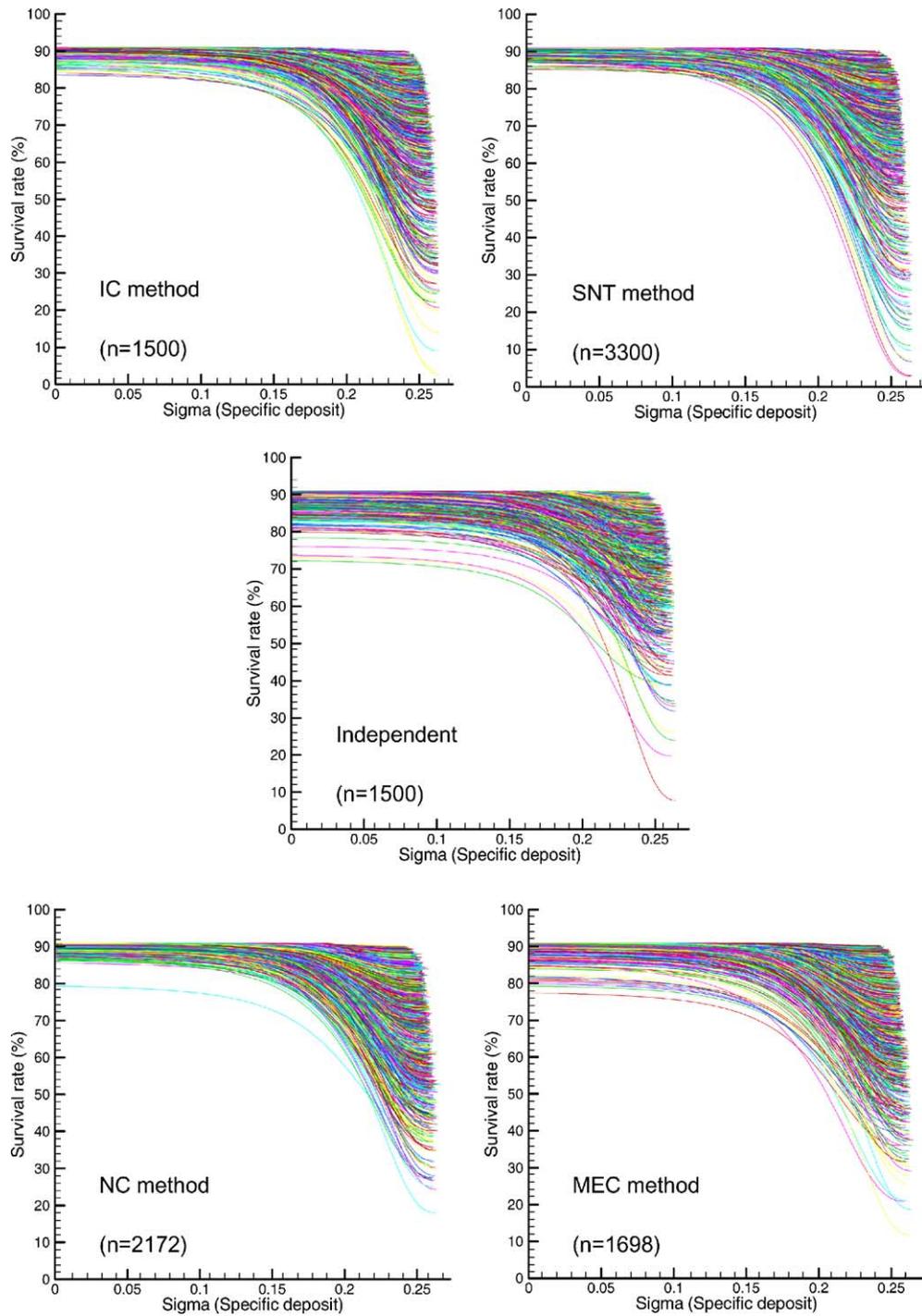


Fig. 5. The outcomes of the second-order MCS using the parameter sets generated with different correlation methods. The spaghetti-looking graph presents  $S$  vs.  $\sigma$  relations corresponding to a number of possible habitat scenarios.

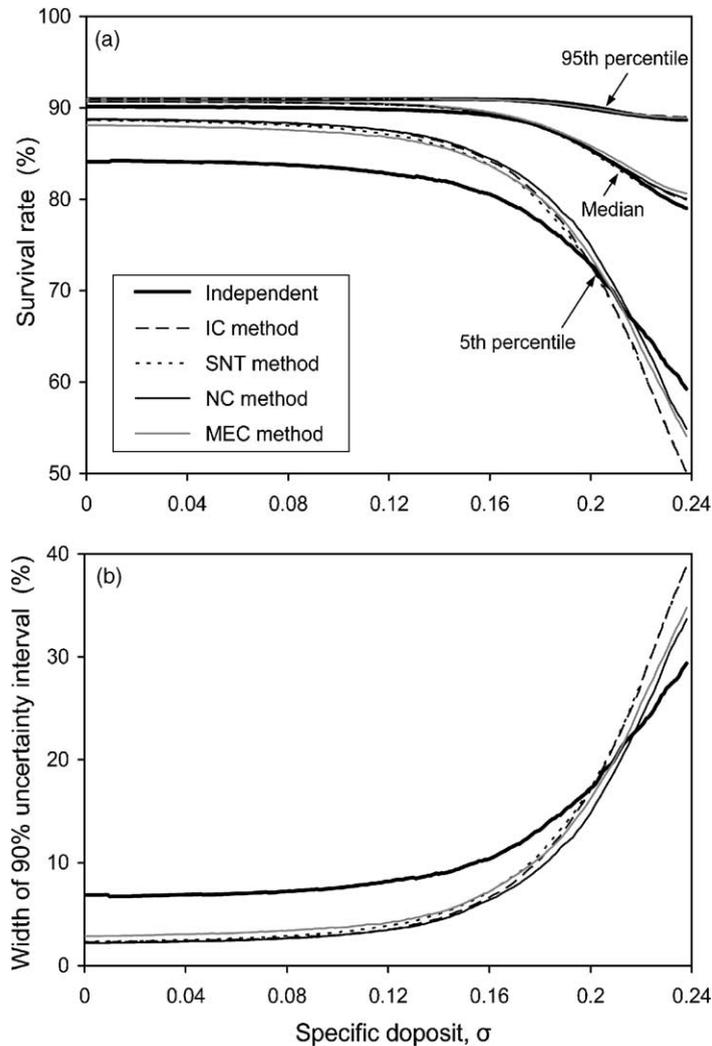


Fig. 6. (a) The 5th and 95th percentiles and median of  $S$  vs.  $\sigma$  relations derived from the spaghetti plots for different correlation methods. (b) The width of the 90% uncertainty intervals (= difference between the 95th and 5th percentile values) for different correlation methods.

### 5.1.2. Variability of embryo survival rate

For any  $S$  versus  $\sigma$  curve in the spaghetti plot (i.e. for any set of model parameters), the variability of embryo survival rate induced by the system variability, denoted as  $\Delta S_v$ , is defined by  $S(0) - S(\sigma_{\text{sat}})$ , where  $S(0)$  and  $S(\sigma_{\text{sat}})$  are the  $S$ -values corresponding to  $\sigma = 0$  and  $\sigma_{\text{sat}}$  (= saturated value), respectively (as illustrated in Fig. 4). Because each  $S$  versus  $\sigma$  curve is associated with a  $\Delta S_v$ -value, the  $\Delta S_v$ -values of all curves in a spaghetti plot are calculated and the histogram is shown in Fig. 7 for the outcomes of different

correlation methods. The survival-rate variability  $\Delta S_v$  ranges between 0 and 90%, mostly between 0 and 20% but rarely beyond 70%. To further examine the distribution of the survival-rate variability, the statistical moments (mean, standard deviation, skewness and kurtosis coefficients) of the  $\Delta S_v$ -values are calculated. From Table 3 we see that the mean values, skewness and kurtosis coefficients are similar for the independent and correlated scenarios, with the independent mean  $\Delta S_v$ -value being 13.7% and the correlated ones ranging from 10.2 to 16.1%. The skewness coefficients

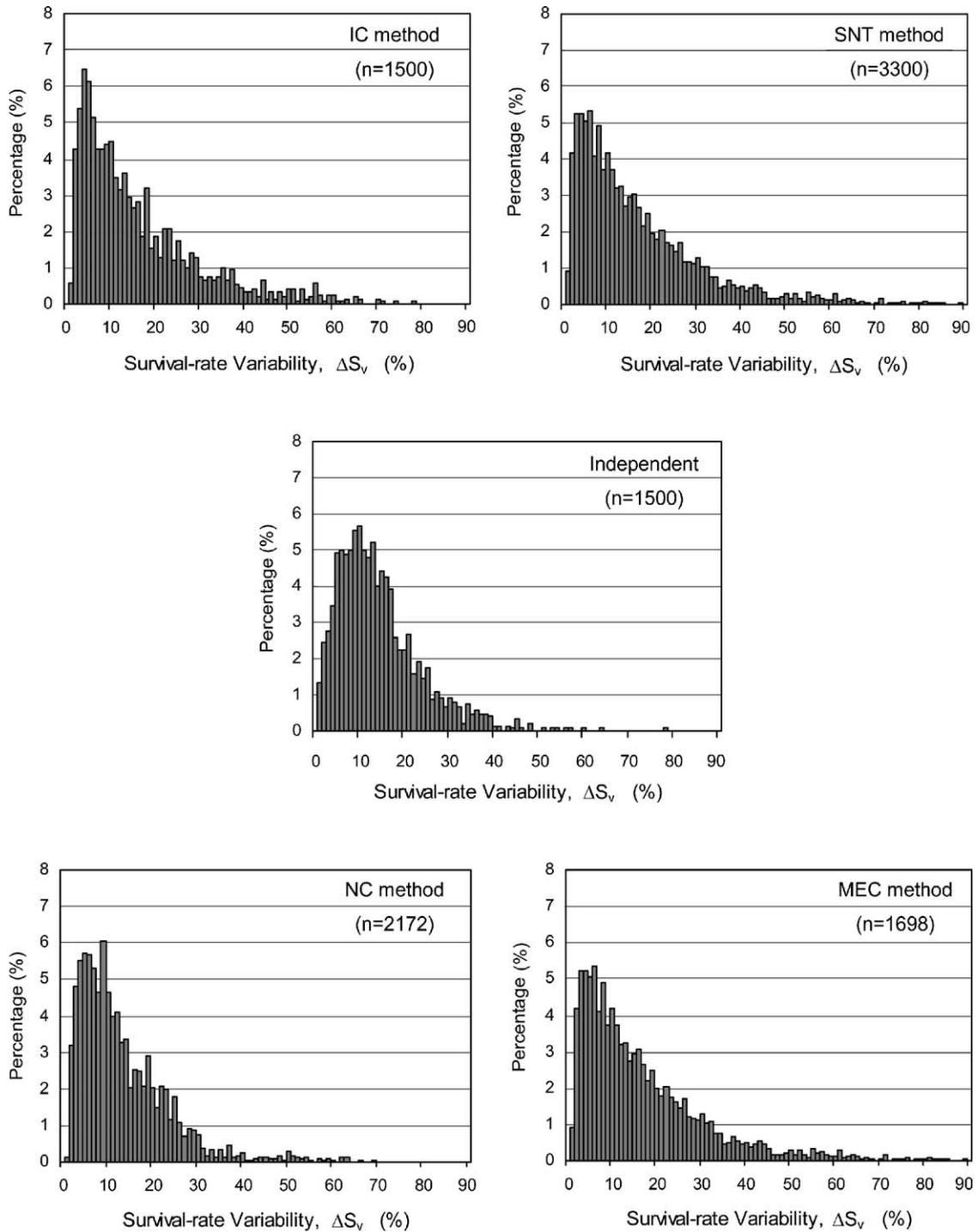


Fig. 7. Histograms of survival-rate variability for different correlation methods. (The increment of survival-rate variability = 1%.)

Table 3  
Statistical moments of survival-rate variability for different methods

Method	Mean (%)	S.D. (%)	Skewness	Kurtosis
Independent	13.7	9.4	1.5	6.6
IC	15.6	13.8	1.6	5.7
SNT	16.1	14.2	1.7	7.5
NC	10.2	12.9	1.8	7.6
MEC	15.0	12.7	1.2	3.9

are all positive (ranging from 1.2 to 1.8) and the kurtosis coefficients vary between 3.9 and 7.6. However, a significant discrepancy exists between the standard deviations for the independent and correlated scenarios. The correlated standard deviations are within a narrow range between 12.7 and 14.2%, whereas the independent standard deviation (=9.4%) is much less than the correlated ones. Given that the magnitude of standard deviation represents the uncertainty in a distribution, the implication of such discrepancy would be that by ignoring the parameter correlations the uncertainty of survival-rate variability is underestimated.

Based on the results presented above, it can be concluded that the outcomes corresponding to the parameters generated with different correlation methods are very similar. However, the IC method can be selected as the simplest and most satisfactory one in view of its efficiency and overall performance. In the subsequent sections, although four correlation methods are used in the analyses, only the outcomes of IC method are used to demonstrate the results associated with the correlated scenario.

## 5.2. Definition of favorable spawning habitat

### 5.2.1. Favorable habitat conditions

It would be natural for the salmon conservationists to ask: what constitute the physical conditions of a favorable habitat? To answer this, herein we define a favorable habitat as the habitat with  $S(\sigma_{\text{sat}}) > 70\%$ . To examine the physical conditions that facilitate the favorable habitats, the model parameters corresponding to the favorable  $S$  versus  $\sigma$  curves are identified. Two findings are gained from these results. First, the distributions of the favorable conditions are obtained, as shown in Fig. 8. For the three parameters, the favorable ranges for the independent scenario are consistently larger than those for the correlated one.

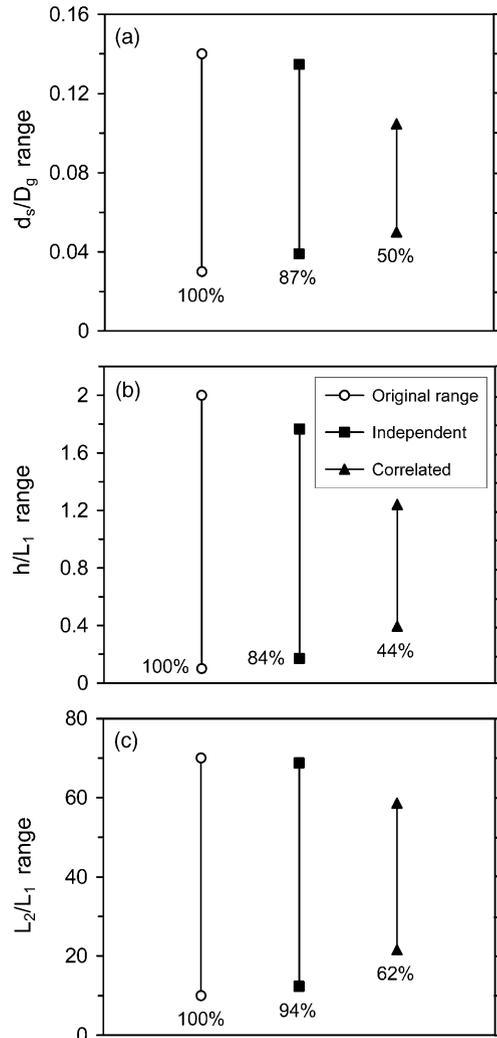
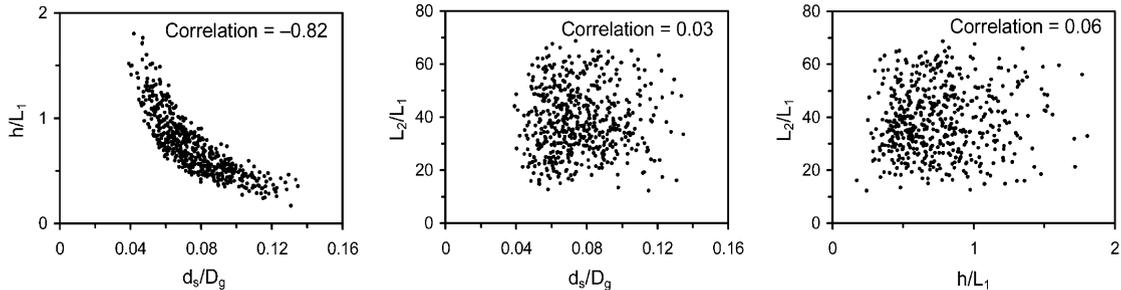


Fig. 8. Distributions of favorable habitat conditions for independent and correlated scenarios. (The original ranges of the three parameters are shown for comparison.)

The independent favorable parameters cover 87, 84, and 94% of the original ranges for  $d_s/D_g$ ,  $h/L_1$ , and  $L_2/L_1$ , respectively, while the corresponding ranges of the correlated favorable parameters only cover 50, 44, and 62% of the original ranges, primarily within the middle-lower portions. These much smaller ranges of the correlated favorable parameters imply that the independent scenario might be overoptimistic, especially for the parameter  $h/L_1$ . The favorable range of  $h/L_1$  for the correlated scenario is

## (a) Independent



## (b) Correlated

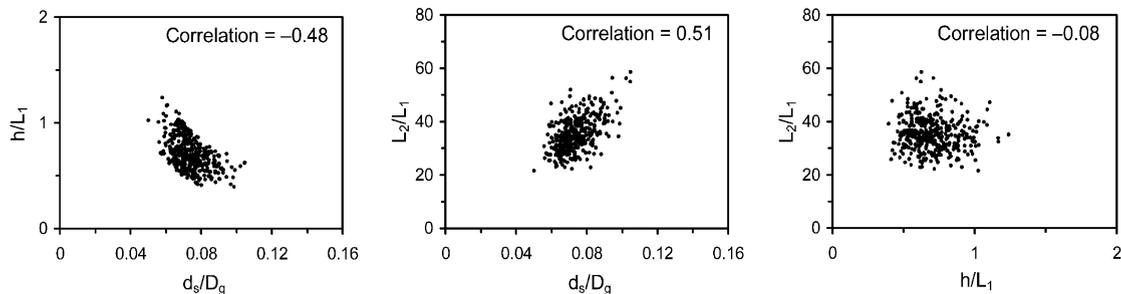


Fig. 9. Dependences among favorable habitat conditions for independent and correlated scenarios.

only half of the corresponding range for the independent one. Much of the favorable range overestimated with the independent scenario is in the uppermost portion.

Second, the dependences among the favorable parameters are observed (Fig. 9). For the correlated scenario, apparent negative and positive dependences are demonstrated in the scatter plots of  $h/L_1$  versus  $d_s/D_g$  and  $L_2/L_1$  versus  $d_s/D_g$ , respectively, whereas no obvious dependences are shown in the scatter plots of  $L_2/L_1$  versus  $h/L_1$ . For the independent scenario, an apparent negative dependence is revealed only in the scatter plot of  $h/L_1$  versus  $d_s/D_g$ . To make quantitative comparisons, the correlation coefficients between the favorable parameters are shown in Fig. 9. The correlation coefficient between  $h/L_1$  and  $d_s/D_g$  is  $-0.48$  for the correlated scenario, while the corresponding value for the independent scenario is  $-0.82$ . The physical interpretation of this negative correlation is that the finer sand deposit needs a greater pressure head to facilitate favorable seepage velocity and thus the survival rate. From a management perspective, the much higher negative correlation associated with the inde-

pendent scenario would potentially lead to an overestimated cost for habitat improvement. It is also interesting to note that the dependence between  $h/L_1$  and  $d_s/D_g$  is demonstrated by a band-shaped scatter plot rather than a natural-looking pattern as shown in the scatter plot of  $L_2/L_1$  versus  $d_s/D_g$ , implying that the pressure head is more dominant than the length of flow path in affecting the seepage velocity and thus the survival rate for a given  $d_s/D_g$ . On the other hand, the correlation between  $L_2/L_1$  and  $d_s/D_g$  is  $0.51$  for the correlated scenario, whereas no obvious correlation is observed for the independent one. The physical interpretation of this positive correlation is that the finer sand deposit needs a shorter flow path to facilitate the favorable seepage velocity and thus the survival rate. The much lower correlation associated with the independent scenario would probably lead to an ineffective design of habitat improvement measures. In general, the correlations between  $L_2/L_1$  and  $h/L_1$  are low. Such negligible correlations imply that the pressure head and the length of flow path are unlikely to constrain each other in the formation of favorable habitats.

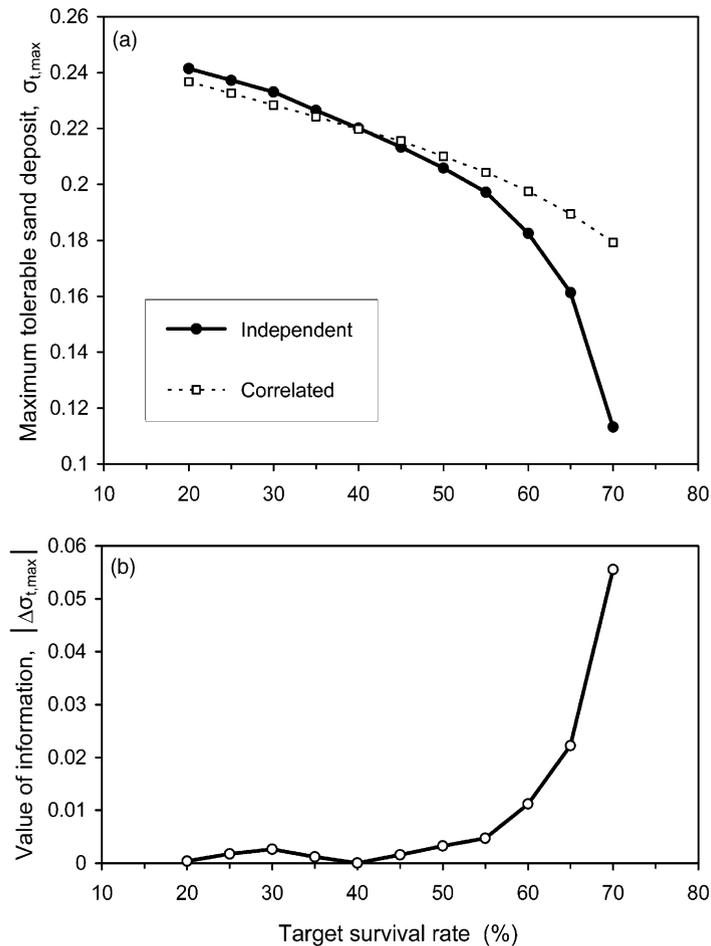


Fig. 10. (a) Maximum tolerable sand deposit vs. target survival rate for independent and correlated scenarios. (b) Value of information vs. target survival rate.

### 5.2.2. Maximum tolerable sand deposit for target survival rate

To implement a project aimed to improve the spawning habitats for embryo development, it would be necessary to set a target survival rate. Once the target is set, the mitigating measures are taken to control the deposition of fine sand into the spawning gravels. What must be known at the planning stage is the amount of sand deposit tolerable for achieving the target survival rate. The spaghetti plots given in Fig. 5 may be used to determine the maximum tolerable sand deposit corresponding to the specified target. For a specified target survival rate (represented by a horizontal line in the spaghetti plot), the maximum tolerable

specific deposit, denoted as  $\sigma_{t,max}$ , is determined by the first intersection (from the left side) with  $S$  versus  $\sigma$  curves. As each  $S$  versus  $\sigma$  curve corresponds to a possible habitat scenario, this intersection represents a most conservative estimate of  $\sigma$  (and thus a maximum tolerable  $\sigma$ -value) for a least favorable combination of habitat conditions. The  $\sigma_{t,max}$ -values corresponding to various target survival rates are shown in Fig. 10a. The  $\sigma_{t,max}$ -values consistently decrease with target survival rate for both scenarios. The independent  $\sigma_{t,max}$ -values become smaller than the correlated ones as the specified target survival rates are higher than 40%. Since the habitat improvement projects are most likely aimed at those higher target values, the

underestimated  $\sigma_{t,\max}$ -values associated with the independent scenario would probably lead to overestimated costs.

To further demonstrate the value of information associated with the parameter correlations, the absolute difference between the correlated and independent  $\sigma_{t,\max}$ -values, denoted as  $|\Delta\sigma_{t,\max}|$ , is plotted against the target survival rate. It is shown in Fig. 10 b that the value of information increases with target survival rate (primarily at the higher target values), indicating that incorporating the parameter correlations in risk

assessments is crucial for the planning of habitat improvement, especially when the target survival rate is set to a higher standard.

5.2.3. Risk for not achieving target survival rate

Given a target survival rate, the corresponding risk for not achieving the specified target is of important concern to risk assessors and decision makers. Such a risk, arising from the knowledge uncertainty in spawning habitats and natural variability in sand deposits, may be evaluated with the spaghetti plots given in

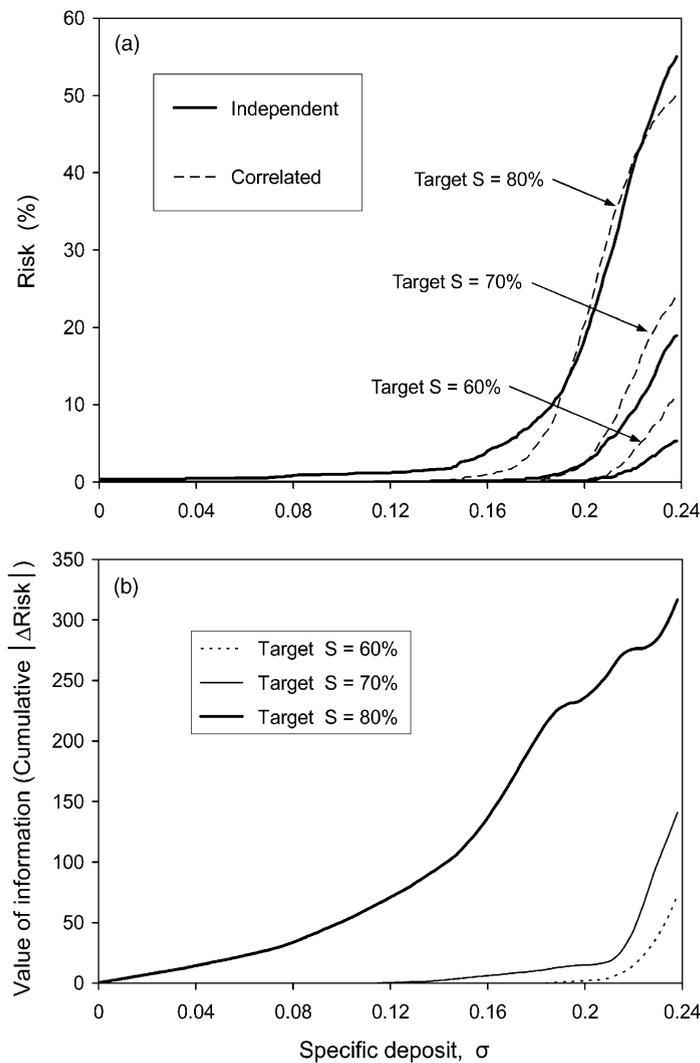


Fig. 11. (a) Variations of risk for independent and correlated scenarios. The risk for not achieving the target survival rate varies with the specific deposit. (b) Value of information vs. specific deposit for different target survival rates.

**Fig. 5.** For a target survival rate, the risk at some  $\sigma$ -value is defined as the percentage of the samples for which the corresponding  $S$ -values are below the specified target. The resulting risk diagram, expressed as a function of target survival rate and specific deposit, is shown in Fig. 11a. The risks associated with different correlation scenarios consistently increase with specific deposit. Significant rises of risk occur at  $\sigma > 0.16, 0.18,$  and  $0.2$  for the target  $S = 80, 70,$  and  $60\%$ , respectively. The risks corresponding to the higher target survival rate are consistently greater than those corresponding to the lower one. No consistent trend between the independent and correlated risks is observed.

To assess the value of information associated with the parameter correlations, the absolute difference between the correlated and independent risks, denoted as  $|\Delta\text{Risk}|$ , is calculated and its cumulative value is plotted against  $\sigma$  in Fig. 11b. For the lower two targets (60 and 70%), significant values of information are associated with those worst specific sand deposits ( $\sigma > 0.2$ ). However, the value of information corresponding to the highest target (80%) is much greater than those corresponding to the lower ones, consistent with the result obtained previously.

## 6. Summary and conclusions

This paper addresses two important issues involved in the survival risk assessment. First, a second-order MCS is employed to separately propagate the knowledge uncertainty and natural variability of the system. Second, the effect of parameter correlations and the associated value of information are investigated. Four multivariate methods are used to generate correlated model parameters. The outcome of the second-order MCS, called “spaghetti plot,” is used to assess the uncertainty, variability, and risk of salmonid embryo survival. The following conclusions can be drawn from this study.

(1) Despite the extreme sampling effect demonstrated in the spaghetti plot, the 90% uncertainty intervals are consistent for all correlated scenarios and thus provide a commendable means to analyze the survival-rate uncertainty. The survival-rate uncer-

tainty increases with the amount of sand deposit. The survival-rate uncertainty associated with the independent scenario is overestimated for the lower values of specific sand deposit but underestimated for those worst sand-deposit conditions.

- (2) The distribution of the survival-rate variability is positively skewed, mostly ranging between 0 and 20%. The uncertainty of the survival-rate variability associated with the independent scenario is underestimated.
- (3) The ranges of favorable habitat conditions for the correlated scenario are significantly smaller than the corresponding ranges for the independent one. The favorable pressure-head and grain-size ratios are negatively correlated; the favorable grain-size and flow-path ratios are positively correlated, whereas no apparent correlation is observed between the favorable pressure-head and flow-path ratios. The independent scenario overestimates the negative correlation between the favorable pressure-head and grain-size ratios, but underestimates the positive correlation between the favorable grain-size and flow-path ratios, both of which would lead to ineffective planning of habitat improvement measures.
- (4) The maximum tolerable sand deposit for achieving the target survival rate decreases with the increasing target value. The results for the independent scenario are more conservative at those higher target values. The risk for not achieving the target survival rate increases with the specified target value and the amount of sand deposit. Significant values of information are associated with the parameter correlations, especially when the higher levels of embryo survival are targeted.
- (5) The scatter plots of the random numbers generated with the four methods demonstrate different correlation patterns. An important finding of this study is that the outputs of various convolutions involved in the uncertainty/variability analyses and risk assessments are consistent for all correlated scenarios, regardless of the diverse correlation patterns. The outcomes associated with the independent scenario, however, substantially differ from those of the correlated ones, indicating that parameter correlations should be incorporated into the Monte Carlo risk assessments whenever possible.

In this paper we establish a paradigm for uncertainty/variability analyses in the ecological context. Presently a systematic framework for such analyses is not available, thus the procedure proposed in this work may provide useful techniques to ecological risk assessors. More field data should be collected in the future, and alternative PDFs of model parameters may be used to investigate the distribution effect.

## Acknowledgements

This research was supported by the National Science Council, Republic of China through Ecological Engineering Grant. We appreciate comments from two anonymous reviewers and the editor, which helped improve the clarity of the paper.

## Appendix A.

### A.1. Iman–Conover method (IC method)

The procedure presented by Iman and Conover (1982) is the most well known method for generating correlated random numbers. Existing risk assessment software, such as Crystal Ball or @Risk, uses IC method to produce correlated random numbers with any set of marginal distributions and specified values of rank–order correlation. The theoretical basis of IC method is that the independent random numbers can be transformed into the correlated ones using the orthogonal transformation. Given  $N$  sets of  $k$  independent random numbers  $(X_1, \dots, X_k)$ , expressed by a  $N \times k$  matrix  $\mathbf{X}$ , one can transform them into the correlated random numbers corresponding to a given rank–order correlation matrix  $\mathbf{R}^*$  by  $\mathbf{X}^* = \mathbf{X} \cdot \mathbf{L}^T$ , where  $\mathbf{X}^*$  is a  $N \times k$  matrix of correlated random numbers  $X_1^*, \dots, X_k^*$ ,  $\mathbf{L}$  is the lower triangular matrix of  $\mathbf{R}^*$ , obtained by Cholesky decomposition  $\mathbf{R}^* = \mathbf{L} \cdot \mathbf{L}^T$ . Generation of correlated random numbers by IC method also involves the use of van der Waerden scores  $a(i)$ , defined as  $a(i) = \Phi^{-1}(i/(N+1))$  for  $i = 1, \dots, N$ , where  $\Phi^{-1}$  is the inverse SN CDF. The stepwise procedure of IC method is as follows:

- *Step 1:* Generate a  $N \times k$  random score matrix  $\mathbf{A}$ , each row of which consists of  $k$  independent com-

ponents randomly sampled from van der Waerden scores.

- *Step 2:* Transform the independent score matrix  $\mathbf{A}$  into the correlated score matrix  $\mathbf{A}^*$  by  $\mathbf{A}^* = \mathbf{A} \cdot \mathbf{L}^T$ . Then the row components of  $\mathbf{A}^*$  have a rank–order correlation matrix  $\mathbf{M}$  close to the given  $\mathbf{R}^*$ .
- *Step 3:* Generate a  $N \times k$  matrix  $\mathbf{X}$  of independent random numbers  $X_1, \dots, X_k$ , each column of which contains  $N$  random numbers sampled from the corresponding marginal distribution.
- *Step 4:* Rearrange the components in each column of  $\mathbf{X}$  such that they have the same rank order as the corresponding column of  $\mathbf{A}^*$ . The resulting matrix  $\mathbf{X}^*$  contains  $N$  sets of random numbers  $X_1^*, \dots, X_k^*$ , which also have a rank–order correlation matrix  $\mathbf{M}$  close to  $\mathbf{R}^*$ .
- *Step 5:* Check if the means and standard deviations of the generated random numbers  $X_1^*, \dots, X_k^*$  converge to the prescribed values. Throughout this study, the convergence criterion is set as  $|\varepsilon| \leq 0.015$ , where  $\varepsilon$  is the difference between the specified and sample values. If the convergence test is passed, proceed to step 6. Otherwise, reject the random numbers, and return to step 1 with  $N = N + 100$ .
- *Step 6:* Use Kolmogorov–Smirnov (K–S) test to examine if the generated random numbers preserve the given marginal distributions. Throughout this study, a significance level of 0.05 is used in the K–S test. If the goodness-of-fit test is passed, proceed to step 7. Otherwise, reject the generated random numbers, and return to step 1 with  $N = N + 100$ .
- *Step 7:* Check the correlations among the generated random numbers. Throughout this study, the acceptance criterion is set as  $|\varepsilon_c| \leq 0.05$ , where  $\varepsilon_c$  is the difference between the specified and observed values of correlation. If the correlation test is passed, then stop. Otherwise, reject the generated random numbers, and return to step 1 with  $N = N + 100$ .

### A.2. Standard normal transformation method (SNT method)

The SNT method is based on the transformation of random variates between the SN and marginal probability spaces (Der Kiureghian and Liu, 1986; Todd and Ng, 2001). For two correlated random numbers  $X_i$  and  $X_j$  in the marginal space, their product–moment

correlation coefficient,  $r_{ij}$ , is related to the equivalent correlation coefficient in the SN space,  $r_{N,ij}$ , as the following:

$$r_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left( \frac{x_i - \mu_i}{\sigma_i} \right) \left( \frac{x_j - \mu_j}{\sigma_j} \right) \times \phi_{ij}(y_i, y_j, r_{N,ij}) \, dy_i \, dy_j \tag{A.1}$$

where  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of  $X_i$ , respectively;  $\phi_{ij}$  is the bivariate SN joint PDF;  $Y_i$  and  $Y_j$  are two correlated SN random variates, obtained by the SN inverse transformation  $Y_i = \Phi^{-1}(F_i(X_i))$ , in which  $F_i$  is the marginal CDF of  $X_i$ . The joint PDF  $\phi_{ij}$  can be expressed as:

$$\begin{aligned} &\phi_{ij}(y_i, y_j, r_{N,ij}) \\ &= \frac{1}{2\pi\sqrt{1 - r_{N,ij}^2}} \exp \left[ -\frac{y_i^2 - 2r_{N,ij}y_iy_j + y_j^2}{2(1 - r_{N,ij}^2)} \right] \end{aligned} \tag{A.2}$$

for  $-1 < r_{N,ij} < 1$ . Given the correlation coefficient  $r_{ij}$  and marginal distributions, the equivalent correlation coefficient  $r_{N,ij}$  can be solved using Eqs. (A.1) and (A.2), which can be expressed as a simplified semi-empirical form (Der Kiureghian and Liu, 1986):

$$r_{N,ij} = T_{ij} \cdot r_{ij} \tag{A.3}$$

where  $T_{ij}$  is the transformation factor. For the correlations and marginals given in Section 2.2, the values of  $T_{12}$ ,  $T_{13}$ , and  $T_{23}$  so obtained are 1.002, 1.004, and 1.002, respectively, in which the subscripts 1, 2, and 3 correspond to the parameters  $d_s/D_g$ ,  $h/L_1$ , and  $L_2/L_1$ , respectively. The  $T_{ij}$ -values given above are close to unity, indicating that the differences between  $r_{ij}$  and  $r_{N,ij}$  are negligible in our setting. Generation of correlated random numbers with SNT method is proceeded as follows:

- *Step 1:* Transform the given product-moment correlation matrix  $\mathbf{R}$  between marginals into the correlation matrix  $\mathbf{R}_N$  between SN using Eq. (A.3).
- *Step 2:* Generate  $N$  sets of independent SN random numbers  $(z_1, \dots, z_k)$ .
- *Step 3:* Transform the independent SN random variates  $Z_i$  into the correlated SN random variates  $Y_i$  by orthogonal transformation  $\mathbf{Y} = \mathbf{L} \cdot \mathbf{Z}$ , where  $\mathbf{L}$  is the lower triangular matrix of  $\mathbf{R}_N$ , obtained by Cholesky decomposition  $\mathbf{R}_N = \mathbf{L} \cdot \mathbf{L}^T$ .

- *Step 4:* Transform the correlated SN random variates  $Y_i$  into the correlated marginal random variates  $X_i$ , using the inverse transformation  $X_i = F_i^{-1}[\Phi(Y_i)]$ .
- *Step 5:* Check the convergence of the means and standard deviations of the generated random numbers  $X_i$ . If the test is passed, proceed to step 6. Otherwise, reject the generated random numbers and return to step 2 with  $N = N + 100$ .
- *Step 6:* Use K–S test to examine the goodness of fit. If passed, proceed to step 7. Otherwise, reject the generated random numbers and return to step 2 with  $N = N + 100$ .
- *Step 7:* Check the correlations  $r_{ij}$  among the generated random numbers. If the correlation test is passed, then stop. Otherwise, reject the generated random numbers and return to step 2 with  $N = N + 100$ .

### A.3. Normal copula method (NC method)

The multivariate normal copula has been used in the decision and risk analysis (Clemen and Reilly, 1999). Like other copula families, the multivariate normal copula allows any marginal distributions. In addition, the normal copula permits the use of multivariate pairwise correlations. The flexibility and analytical tractability of the multivariate normal copula make it an attractive method for generating correlated random numbers. The basic idea of a copula is that the joint distribution of random variates can be expressed as a function of the marginal distributions (Sklar, 1959), i.e.

$$F(x_1, \dots, x_k) = C[F_1(x_1), \dots, F_k(x_k)] \tag{A.4}$$

where  $F(x_1, \dots, x_k)$  is the joint CDF for random variates  $X_1, \dots, X_k$  with marginal CDFs  $F_1(x_1), \dots, F_k(x_k)$ ; and  $C$ , called a copula, is a joint CDF with uniform marginals on  $[0,1]$ . Given that  $F_i(x_i)$  and  $C$  are both differentiable, the joint PDF  $f(x_1, \dots, x_k)$  can be expressed as:

$$\begin{aligned} f(x_1, \dots, x_k) &= f_1(x_1) \times \dots \times f_k(x_k) \\ &\times c[F_1(x_1), \dots, F_k(x_k)] \end{aligned} \tag{A.5}$$

where  $f_i(x_i)$  is the marginal PDF corresponding to  $F_i(x_i)$ , and  $c = \partial^k C / (\partial F_1 \dots \partial F_k)$  is the copula density. Eq. (A.5) reveals that the joint PDF is a product of the marginal PDFs and the copula density.

For independent  $X_i$ 's,  $c = 1$  and  $f(x_1, \dots, x_k) = f_1(x_1) \times \dots \times f_k(x_k)$ , which implies that  $c$  encodes information about the dependence (or correlation) among the  $X_i$ 's. Hence,  $c$  is also called a dependence function.

The multivariate normal distribution is typically parameterized with a matrix  $\mathbf{R}_N$  of product-moment correlation coefficients  $r_{N,ij}$ , while the multivariate normal copula is defined by a rank-order correlation matrix  $\mathbf{R}^*$  (Clemen and Reilly, 1999). Solving Eq. (A.5) for the normal copula density  $c_N$  in terms of the multivariate SN joint PDF  $\phi(y_1, \dots, y_k | \mathbf{R}_N)$  yields:

$$c_N[\Phi(y_1), \dots, \Phi(y_k) | \mathbf{R}^*] = \frac{\phi(y_1, \dots, y_k | \mathbf{R}_N)}{[\phi(y_1) \times \dots \times \phi(y_k)]} \tag{A.6}$$

where  $\Phi(Y_i)$  and  $\phi(Y_i)$  are the univariate SN CDF and PDF, respectively, in which  $Y_i$  is obtained by  $Y_i = \Phi^{-1}[F_i(X_i)]$ . Substituting the SN PDFs into Eq. (A.6) and using the linear algebra, Clemen and Reilly (1999) obtained an expression for  $c_N$ :

$$c_N[\Phi(y_1), \dots, \Phi(y_k) | \mathbf{R}^*] = \frac{\exp(-\mathbf{y}^T(\mathbf{R}_N^{-1} - \mathbf{I})\mathbf{y}/2)}{|\mathbf{R}_N|^{1/2}} \tag{A.7}$$

where  $\mathbf{y} = (y_1, \dots, y_k)^T$ ,  $\mathbf{I}$  is the  $k \times k$  identity matrix, and  $|\mathbf{R}_N|$  is the determinant of  $\mathbf{R}_N$ . The normal copula density given in Eq. (A.7) can be used to generate correlated random numbers by an acceptance/rejection procedure (Gentle, 1998; Haas, 1999), as described below:

- *Step 1:* Transform the given product-moment correlation matrix  $\mathbf{R}$  between marginals into the correlation matrix  $\mathbf{R}_N$  between SN using Eq. (A.3).
- *Step 2:* Determine the maximum  $c_N$ -value on the  $k$ -dimensional unit hypercube, and call this  $c_{N,\max}$ , where  $c_N$  is calculated with Eq. (A.7).
- *Step 3:* Draw  $N$  sets of uniform random numbers  $(u_1, \dots, u_k, \xi)$  from  $[0,1]$ .
- *Step 4:* For each set, if  $\xi \leq c_N(u_1, \dots, u_k)/c_{N,\max}$ , then accept  $(u_1, \dots, u_k)$ ; otherwise reject this set, where  $c_N(u_1, \dots, u_k)$  is evaluated with Eq. (A.7). The accepted uniform random numbers are transformed to the marginal using  $X_i = F_i^{-1}(U_i)$ .

- *Step 5:* Check if the means and standard deviations of the generated random numbers  $X_i$  converge to the prescribed values. If the test is passed, proceed to step 6. Otherwise, reject the generated random numbers and return to step 3 with  $N = N + 100$ .
- *Step 6:* Use K-S test to examine the goodness of fit. If passed, proceed to step 7. Otherwise, reject the generated random numbers and return to step 3 with  $N = N + 100$ .
- *Step 7:* Check the correlations  $r_{ij}$  among the generated random numbers. If the correlation test is passed, then stop. Otherwise, reject the generated random numbers and return to step 3 with  $N = N + 100$ .

#### A.4. Maximum-entropy copula method (MEC method)

The piecewise-uniform copula presented by MacKenzie (1994) is applicable to generating correlated random numbers. Determination of the piecewise-uniform copula density is based on the principle of entropy maximization. Shannon (1949) proposed a measure of the expected amount of information contained in an observed probability distribution, which is called entropy and defined as the following for a discrete case:

$$\text{entropy} = -\sum_{i=1}^k p_i \ln(p_i) \tag{A.8}$$

where  $p_1, \dots, p_k$  are the probabilities corresponding to a set of events  $\{E_1, \dots, E_k\}$  such that  $\sum p_i = 1$ . For a set of random numbers  $\{X_1, \dots, X_k\}$  having a joint PDF  $f(x_1, \dots, x_k)$ , the entropy is defined as:

$$\text{entropy}(f) = -\int \dots \int \ln(f(x_1, \dots, x_k)) \cdot f(x_1, \dots, x_k) dx_1 \dots dx_k \tag{A.9}$$

Maximizing entropy subject to the given constraints is considered by many to be an intuitively compelling paradigm. By doing so one would be maximizing the amount of information to be learned from observing a manifestation from  $f$ . Based on Eq. (A.5) and the definition of entropy given by Eq. (A.9), the following relation is obtained (MacKenzie, 1994), which expresses the entropy of the joint PDF with the entropies of the marginals and of

the copula, i.e.

$$\text{entropy}(f) = \sum_{i=1}^k \text{entropy}(f_i) + \text{entropy}(c) \quad (\text{A.10})$$

where

$$\text{entropy}(c) = - \int \cdots \int \ln(c(u_1, \dots, u_k)) \times c(u_1, \dots, u_k) du_1 \cdots du_k \quad (\text{A.11})$$

in which  $c(u_1, \dots, u_k)$  is the joint copula density. As revealed by Eq. (A.10), if the marginal distributions are known, then the problem of maximizing entropy reduces to the problem of finding the maximum-entropy copula subject to the given constraints on dependence. The piecewise-uniform copula provides a useful tool for solving such a problem.

The  $k$ -dimensional piecewise-uniform copula density  $c_{k,m}(u_1, \dots, u_k)$  is constructed by partitioning each marginal on the unit hypercube into  $m$  equal-size pieces. The resulting  $m^k$  combinations of intervals are denoted as  $I_{i_1, \dots, i_k}$ , where the subscript  $i_j$  indicates that the uniform random variable  $U_j$  belongs to the interval  $[(i_j - 1)/m, i_j/m]$ , for  $i_j = 1, \dots, m$  and  $j = 1, \dots, k$ . Each  $I_{i_1, \dots, i_k}$  is associated with a constant density, as expressed by:

$$c_{k,m}(u_1, \dots, u_k) = p_{i_1, \dots, i_k} \quad (\text{A.12})$$

The piecewise-uniform copula density  $p_{i_1, \dots, i_k}$  is subject to the following constraints:

$$p_{i_1, \dots, i_k} \geq 0 \quad (\text{A.13})$$

and

$$\frac{p_{i, \dots}}{m^k} = \frac{p_{\dots, i, \dots}}{m^k} = \frac{p_{\dots, i}}{m^k} = \frac{1}{m} \quad \text{for } i = 1, \dots, m \quad (\text{A.14})$$

where  $p_{\dots}$  denotes the summation of piecewise copula densities over the remaining  $m^{k-1}$  combinations of subscript. Eqs. (A.13) and (A.14) are to assure that  $p_{i_1, \dots, i_k}$  is a copula density and has uniform marginals. An additional constraint is provided to model the dependence among the  $k(k - 1)/2$  pairs of random variates, which is given by the following formula for Spearman's  $\rho$  in terms of the copula density (Nelsen, 1999). For the multivariate case, the correlation coefficient  $\rho_{jl}$  between two random variates  $U_j$  and  $U_l$  can

be expressed as:

$$\rho_{jl} = 12 \int_0^1 \cdots \int_0^1 u_j \cdot u_l \times c(u_1, \dots, u_j, \dots, u_l, \dots, u_k) du_1 \cdots du_k - 3 \quad (\text{A.15})$$

Using the piecewise-uniform copula density, we turn Eq. (A.15) into the following form:

$$\begin{aligned} \rho_{jl} &= 12 \int_0^1 \int_0^1 u_j \cdot u_l \left[ \sum_{i_1=1}^m \cdots \sum_{i_k=1}^m p_{i_1, \dots, i_j, \dots, i_l, \dots, i_k} \right. \\ &\quad \left. \times \int_{(i_1-1)/m}^{i_1/m} \cdots \int_{(i_k-1)/m}^{i_k/m} du_1 \cdots du_k \right] \times du_j du_l - 3 \\ &= \frac{3}{m^{k+2}} \left\{ \sum_{i_j=1}^m \sum_{i_l=1}^m \left[ \sum_{i_1=1}^m \cdots \sum_{i_k=1}^m p_{i_1, \dots, i_j, \dots, i_l, \dots, i_k} \right] \right. \\ &\quad \left. \times (2i_j - 1)(2i_l - 1) \right\} - 3 \quad (\text{A.16}) \end{aligned}$$

Moreover, with the piecewise-uniform copula density, Eq. (A.11) can be rewritten as:

$$\text{entropy}(c) = - \frac{1}{m^k} \sum_{i_1=1}^m \cdots \sum_{i_k=1}^m \ln(p_{i_1, \dots, i_k}) \cdot p_{i_1, \dots, i_k} \quad (\text{A.17})$$

MacKenzie (1994) has shown that the optimization problem defined by maximizing Eq. (A.17) subject to the constraints given by Eqs. (A.13), (A.14) and (A.16) has a unique global solution. The  $m^k$  piecewise-uniform copula densities  $p_{i_1, \dots, i_k}$  so determined form a basis for generating correlated random numbers. Before we proceed to the generation procedure, it would be appropriate to give a criterion here for determining the minimum value of  $m$ . It is shown that  $\pm(1 - 1/m^2)$  are the limiting values of pairwise correlation captured by a piecewise-uniform copula (MacKenzie, 1994), thus  $m$  must be large enough to assure that all  $|\rho_{jl}| \leq 1 - 1/m^2$ . For the correlations given in Table 2,  $m = 4$  may be taken. The acceptance/rejection procedure used in the NC method is also adopted here for generating correlated random numbers.

- *Step 1:* Given the rank–order correlation matrix  $\mathbf{R}^*$  among the random variates, solve the optimization problem defined by maximizing the entropy given in Eq. (A.17) subject to the constraints given in Eqs. (A.13), (A.14) and (A.16).
- *Step 2:* Find the maximum  $p_{i_1, \dots, i_k}$ -value obtained from step 1, and call this  $c_{k,m,\max}$ .
- *Step 3:* Draw  $N$  sets of uniform random numbers  $(u_1, \dots, u_k, \xi)$  from  $[0,1]$ .
- *Step 4:* For each set, if  $\xi \leq c_{k,m}(u_1, \dots, u_k) / c_{k,m,\max}$ , then accept  $(u_1, \dots, u_k)$ ; or else reject this set, where  $c_{k,m}(u_1, \dots, u_k)$  is the  $p_{i_1, \dots, i_k}$  obtained from step 1.
- *Step 5:* Check the correlations  $\rho_{jl}$  among the accepted uniform random numbers. If the correlation test is passed, proceed to step 6. Otherwise, reject the uniform random numbers and return to step 3 with  $N = N + 100$ .
- *Step 6:* The accepted uniform random numbers are then transformed into the marginal random numbers using  $X_j = F_j^{-1}(U_j)$ .
- *Step 7:* Check the convergence of the means and standard deviations of the generated random numbers  $X_j$ . If the test is passed, proceed to step 8. Otherwise, reject these random numbers and return to step 3 with  $N = N + 100$ .
- *Step 8:* Use K–S test to examine the goodness of fit. If passed, then stop. Otherwise, reject the generated random numbers and return to step 3 with  $N = N + 100$ .

## References

- Anderson, E.L., Hattis, D., 1999. Uncertainty and variability. *Risk Anal.* 19, 47–49.
- Annan, J.D., 2001. Modelling under uncertainty: Monte Carlo methods for temporally varying parameters. *Ecol. Model.* 136, 297–302.
- Bogen, K.T., Spear, R.C., 1987. Integrating uncertainty and interindividual variability in environmental risk assessment. *Risk Anal.* 7, 427–436.
- Bukowski, J., Korn, L., Wartenberg, D., 1995. Correlated inputs in quantitative risk assessment: the effects of distributional shape. *Risk Anal.* 15, 215–219.
- Burmester, D.E., Anderson, P.D., 1994. Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessment. *Risk Anal.* 14, 477–481.
- Burmester, D.E., Wilson, A.M., 1996. An introduction to second-order random variables in human health risk assessments. *Hum. Ecol. Risk Assess.* 2, 892–919.
- Clemen, R.T., Reilly, T., 1999. Correlations and copulas for decision and risk analysis. *Manag. Sci.* 45, 208–224.
- Cooper, A.C., 1965. The effect of transported stream sediments on survival of sockeye and pink salmon eggs and alevins. In: *Int. Pac. Salmon Comm. Bull.*, vol. 18. New Westminster, BC.
- Der Kiureghian, A., Liu, P.L., 1986. Structural reliability under incomplete probability information. *J. Eng. Mech.* 112, 85–104.
- Frey, H.C., Rhodes, D.S., 1996. Characterization, simulation and analyzing variability and uncertainty: an illustration of methods using an air toxics emission example. *Hum. Ecol. Risk Assess.* 2, 762–797.
- Gentle, J.E., 1998. *Random Number Generation and Monte Carlo Methods*. Springer, New York.
- Haas, C.N., 1999. On modeling correlated random variables in risk assessment. *Risk Anal.* 19, 1205–1214.
- Harmon, R., Challenor, P., 1997. A Markov chain Monte Carlo method for estimation and assimilation into models. *Ecol. Model.* 101, 41–59.
- Hattis, D., Burmaster, D.E., 1994. Assessment of variability and uncertainty distributions for practical risk analyses. *Risk Anal.* 14, 713–730.
- Hession, W.C., Storm, D.E., Haan, C.T., Burks, S.L., Matlock, M.D., 1996. A watershed-level ecological risk assessment methodology. *Water Resour. Bull.* 32, 1039–1054.
- Hoffman, F.O., Hammonds, J.S., 1994. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Anal.* 14, 707–712.
- Iman, R.L., Conover, W.J., 1982. A distribution-free approach to inducing rank order correlation among input variables. *Commun. Stat.—Simul. Comput.* 11, 311–334.
- MacKenzie, G.R., 1994. Approximately maximum-entropy multivariate distributions with specified marginals and pairwise correlations. Ph.D. Dissertation, University of Oregon.
- Mitchell, S.W., Csillag, F., 2001. Assessing the stability and uncertainty of predicted vegetation growth under climatic variability: northern mixed grass prairie. *Ecol. Model.* 139, 101–121.
- Moschandreass, D.J., Karuchit, S., 2002. Scenario-model-parameter: a new method of cumulative risk uncertainty analysis. *Environ. Int.* 28, 247–261.
- Nauta, M.J., 2000. Separation of uncertainty and variability in quantitative microbial risk assessment models. *Int. J. Food Microbiol.* 57, 9–18.
- Neave, H.M., Cunningham, R.B., Norton, T.W., Nix, H.A., 1997. Preliminary evaluation of sampling strategies to estimate the species richness of diurnal, terrestrial birds using Monte Carlo simulation. *Ecol. Model.* 95, 17–27.
- Nelsen, R.B., 1999. *An Introduction to Copulas*. Springer, New York.
- Pohlmann K.F., Hassan A.E., Chapman J.B., 2002. Modeling density-driven flow and radionuclide transport at an underground nuclear test: uncertainty analysis and effect of parameter correlation. *Water Resour. Res.* 38(5), doi:10.1029/2001WR001047.
- Pouillot, R., Albert, I., Cornu, M., Denis, J.B., 2003. Estimation of uncertainty and variability in bacterial growth using Bayesian inference: application to *Listeria monocytogenes*. *Int. J. Food Microbiol.* 81, 87–104.

- Qian, S.S., Stow, C.A., Borsuk, M.E., 2003. On Monte Carlo methods for Bayesian inference. *Ecol. Model.* 159, 269–277.
- Rai, S.N., Krewski, D., 1998. Uncertainty and variability analysis in multiplicative risk models. *Risk Anal.* 18, 37–45.
- Rai, S.N., Krewski, D., Bartlett, S., 1996. A general framework for the analysis of uncertainty and variability in risk assessment. *Hum. Ecol. Risk Assess.* 2, 972–989.
- Salinger, D.H., Anderson, J.J., Hamel, O.S., 2003. A parameter estimation routine for the vitality-based survival model. *Ecol. Model.* 166, 287–294.
- Sanga, R.N., Bartell, S.M., Ponce, R.A., Boischio, A.A.P., Joiris, C.R., Pierce, C.H., Faustman, E.M., 2001. Effects of uncertainties on exposure estimates to methylmercury: a Monte Carlo analysis of exposure biomarkers versus dietary recall estimation. *Risk Anal.* 21, 859–868.
- Shannon, C.E., 1949. *The Mathematical Theory of Communication*. University of Illinois Press, Urbana.
- Sklar, A., 1959. Fonctions de répartition à  $n$  dimensions et leurs marges. *Publications de l'Institut Statistique de l'Université de Paris* 8, 229–231.
- Smith, A.E., Ryan, P.B., Evans, J.S., 1992. The effect of neglecting correlations when propagating uncertainty and estimating the population distribution of risk. *Risk Anal.* 12, 467–474.
- Tappel, P.D., Bjornn, T.C., 1983. A new method of relating size of spawning gravel to salmonid embryo survival. *North Am. J. Fisheries Manag.* 3, 123–135.
- Todd, C.R., Ng, M.P., 2001. Generating unbiased correlated random survival rates for stochastic population models. *Ecol. Model.* 144, 1–11.
- Vose, D., 2000. *Risk Analysis*. John Wiley & Sons, Chichester.
- Waller, L.A., Smith, D., Childs, J.E., Real, L.A., 2003. Monte Carlo assessments of goodness-of-fit for ecological simulation models. *Ecol. Model.* 164, 49–63.
- Wu, F.C., 2000. Modeling embryo survival affected by sediment deposition into salmonid spawning gravels: application to flushing flow prescriptions. *Water Resour. Res.* 36, 1595–1603.
- Wu, F.C., Shen, H.W., 1999. First-order estimation of stochastic parameters of a sediment transport model. *J. Hydraulic Res.* 37, 213–227.