This paper has been submitted for publication. Please do not cite without authors permission

Correlations between parameters in risk models: estimation and propagation of uncertainty by Markov Chain Monte Carlo

A E Ades, G Lu

MRC Health Services Research Collaboration, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, England

Correspondence to: Tony Ades, e-mail t.ades@bristol.ac.uk

corr5-ref-2 September 18, 2002

ABSTRACT

Monte Carlo simulation has become the accepted method for propagating parameter uncertainty through risk models. It is widely appreciated, however, that correlations between input variables must be taken into account if models are to deliver correct assessments of uncertainty in risk. Various two-stage methods have been proposed, which first estimate a correlation structure and then generate Monte Carlo simulations which incorporate this structure while leaving marginal distributions of parameters unchanged. Here we propose a one-stage alternative, in which the correlation structure is estimated from the data directly by Bayesian Markov Chain Monte Carlo methods. Samples from the posterior distribution of the outputs then correctly reflect the correlation between parameters, given the data and the model. Besides its computational simplicity, this approach utilises the available evidence from a wide variety of structures, including incomplete data, correlated and uncorrelated repeat observations. The major advantage of a Bayesian approach is that, rather than assuming the correlation structure is fixed and known, it captures the joint uncertainty induced by the data in all parameters, including variances and covariances, and correctly propagates this through the decision or risk model. These features are illustrated with examples on emissions of dioxin congeners from solid waste incinerators.

KEY WORDS:

Markov Chain Monte Carlo, Monte Carlo simulation, Bayesian methods, correlation, uncertainty analysis, dioxins

1. INTRODUCTION

Monte Carlo simulation is the widely accepted method for propagating uncertainty in input parameters through a decision model.^{(1) (2) (3)} Simulation generates probability distributions for model outputs, and enables analyses to be made of the contribution of different inputs to the uncertainty in the outputs. At the same time, it is recognised that correlations between input variables can have a profound influence on the uncertainty in the outputs, and the importance of "correlation control" is stressed in recent textbooks.^{(4) (5)}

Faced with the need to incorporate input correlation in Monte Carlo simulation, the procedure most frequently adopted in practice is that proposed by Iman and Conover.⁽⁶⁾ This is a restrictive pairing technique, now implemented in the Crystal Ball decision package, which induces a specified rank correlation structure in Monte Carlo simulation while leaving the marginal distributions of the input parameters unchanged. The process involves, in essence, generating samples from the marginal distributions of the inputs and then rearranging the same samples in a way that respects the desired rank correlation structure. More recently, copulas, a class of mathematical representations of bivariate distributions on the unit square have been proposed,^{(7) (8) (5)} This method, again, allows the user to specify the marginal distributions of parameters and a matrix of between-parameter dependence measures such as Spearman's ρ or Kendall's τ , and then generates a joint distribution using a copula relationship. A related approach is the NORTA (Normal To Anything) algorithm.⁽⁹⁾ This begins with a specification of the marginal distributions and the desired correlation matrix, and then generates a multi-variate normal distribution which can be mapped into the required joint distribution.

All these methods are based on correlation structures that are assumed to be fixed and known. These are powerful procedures which are particularly attractive whenever the correlations originate from expert opinion and so cannot be measured directly (see Clemen and Reilly ⁽⁸⁾ for a discussion of elicitation methods). However, we will argue that whenever possible, the correlations should be treated as uncertain, and be estimated jointly with the other parameters from available evidence. This can be achieved in a number of ways, but if Bayesian Markov Chain Monte Carlo methods

are used, not only will samples from the joint posterior distribution appropriately reflect the uncertainties in the data, but these uncertainties can be simultaneously propagated through the risk model. This differs fundamentally from other methods so far discussed, not only because it captures the uncertainty in the correlation, but it also because it captures the dependencies between the posterior correlations, variances, and means that are induced by their joint estimation from data.

We begin with the simple bivariate dataset (Table 1) used by Haas,⁽⁷⁾ itself abstracted from an earlier paper by Hattis and Burmaster.⁽¹⁰⁾ This example concerns emissions of two dioxin congeners from seven solid waste incinerators. These emissions are approximately log-normal and positively correlated. The objective of the analysis is to identify the distribution of the overall toxic equivalency (TEq), a simple linear combination of the two dioxins. We then turn to more complex, and more realistic data structures, with incomplete data, repeated observations, and correlations both within and between plants. In discussion we examine the limitations of the Bayesian MCMC approach, the issue of model uncertainty, and the general question of incorporating evidence on complex functions of model parameters.

2. BAYESIAN MARKOV CHAIN MONTE CARLO

Markov Chain Monte Carlo (MCMC) simulation is a computing framework for Bayesian statistical inference. The theoretical background and practical examples are set out in recent texts.^{(11) (12) (13) (14)} Here we employ WinBUGs 1.4 software.⁽¹⁵⁾ The BUGs website (<u>Bayesian inference Using Gibbs Sampling</u>), <u>www.mrc-</u> <u>bsu.cam.ac.uk/bugs/welcome.shtml</u>, gives free access to the software, manuals, worked examples, tutorial papers, lectures, and further references and contacts.

The aim of Bayesian inference is to provide information on the posterior distribution of model parameters given the data. In MCMC samples are drawn from the posterior distribution, and the mean, standard deviation and percentiles of parameters or functions of parameters can then be examined from several thousands of simulated draws. Gibbs sampling ⁽¹⁶⁾ is a form of MCMC algorithm implemented in WinBUGS in which samples are drawn from the conditional distribution of each parameter given the current values of the other parameters and the data. It can be shown that, given

some initial values, the resulting distributions converge to a stationary distribution that correctly reflects the required posterior distributions.⁽¹⁷⁾ The first few thousands of 'burn in' simulations, while the distributions are not yet stationary, are discarded. In the examples presented here, graphical tests incorporated in WinBUGS suggested that adequate convergence had been achieved within 3000 iterations. The first 10,000 were discarded, and results are based on the subsequent 100,000 samples. This runs in 50 seconds on a PC with a 1.9 GHz processor. Datasets and WinBUGS code for each illustration are available from the author.

3. FRAMEWORK

We explain the concepts behind our approach using as an illustration the original data (Table 1) from Haas⁽⁷⁾ giving emission factors for two dioxins from 7 incinerators. The means, standard deviations, on a natural log scale are also given. The correlation coefficient of the logged data is 0.8254, with a 95% confidence interval (0.191 - 0.973), based on the normalising transformation $z= 0.5 \ln((1+r)(1-r))$.

Table 2 presents the Toxic Equivalent (TEq) percentiles obtained by Haas, assuming the Toxic Equivalent Factors for each dioxin are known to be 0.5. These include the results of restricted pairing⁽⁶⁾ based on the observed rank correlation, and a range of copula relationships estimated from the data by maximum likelihood.

For purposes of comparison, a non-Bayesian alternative approach might be to draw samples from two correlated normal distributions, using the means and covariance structure observed in the data (Table 1), exponentiate each sample, and then take an average to obtain TEq. This MC simulation from correlated log-normals gives a somewhat wider distribution than the methods described by Haas. It is not immediately clear why this is, although Hattis & Burmaster also observed that with approximately lognormal data, the simulated joint distributions generated by the Iman and Conover procedure, which is based on rank correlation, consistently underestimated the original correlations.

Before describing a Bayesian analysis of this dataset, we must first clarify exactly which distribution we wish to sample from. Three possible candidates are:

Distribution of mean TEq. Hattis and Burmaster⁽¹⁰⁾ observe that mean TEq and the uncertainty in the mean given this sample from seven incinerators may provide the appropriate perspective for a national policy maker. They treat this as a subsidiary problem and suggest an approximate method.

Distribution of TEq in the seven incinerators based on known parameters. Earlier analyses^{(7) (10)} were intended to deliver the distribution of TEq from the seven plants included in the sample, assuming that the mean, the between-incinerator variance, and the rank correlation between dioxins are all known to be at their maximum likelihood values.

Predictive distribution of TEq in a new incinerator. Here the seven incinerators are considered to have been sampled from a larger population, and in this paper we adopt the perspective of a decision maker concerned with the distribution of TEq that might be expected in a new incinerator, given that the mean and covariance parameters for the population can only be estimated with uncertainty.

4. ILLUSTRATIONS

4.1 Multi-variate normal, no missing data, no repeated measurements

The data in Table 1 gives the emission factors Y_{ij} for congener, j, j=1..N_j, measured in micrograms/metric ton, in incineration plants, i, i=1...N_i. We assume that $y_{ij} = log(Y_{ij})$ has a multi-variate normal distribution:

$$y_{ij} \sim Mnorm(\mu, P)$$

The normality assumption will be explored later. μ is a vector of N_j means, and P an N_j-by-N_j precision matrix, the inverse of an N_j-by-N_j covariance matrix. There is also a vector of N_j known Toxic Equivalency Factors (TEFs), T. In the Table 1 example above, N_j=2, N_i=7, and T=(0.5,0.5).

We assign minimally informative prior distributions to M and P. The two means are assigned a normal prior with low precision (high variance), and the precision matrix a vague Wishart distribution. Other options are available: readers should consult the examples in the WinBUGS manual or tutorial texts on practical Bayesian MCMC modelling.^{(11) (12) (14)} The posterior distribution of interest is the predictive distribution of TEq in an unknown incinerator from this same multi-variate normal population. We draw a pair of values, one for each dioxin, from an 'eighth' incinerator,

exponentiate them back to the natural scale, and multiply them into their respective

TEF and sum :

$$y_{n+1,j} \sim Mnorm(\mu,P)$$

TEq = sum_j (TEF_j exp(y_{n+1,j}))

The full WinBUGS code is as follows:

```
 \begin{array}{ll} model \\ for (i in 1:8) \{ y[i,1:2] \sim dmnorm(mu[],P[,]) \} \\ for (j in 1:2) \{ mu[j] \sim dnorm(0,.0001) \} \\ P[1:2,1:2] \sim dwish(om[,],2) \\ TEq <- (exp(y[8,1]) + exp(y[8,2]) )/2 \\ \} \end{array} \begin{array}{ll} \# 1. \ multivariate normal \ likelihood \\ \# 2. \ vague \ prior \ for \ means \\ \# 3. \ vague \ prior \ for \ precision \ (see \ data \ list) \\ \# 4. \ TEq \ for \ samples \ from \ 'eighth' \ incinerator \\ \end{array}
```

with the following data list:

```
list(y=structure(.Data=c(1.853168097, 3.987130478,
2.93385687, 4.032469159,
0.875468737, 1.289232648,
-1.482805262, 2.4765384,
5.638354669, 5.247024072,
-0.410980289, 1.682688374,
0.506817602, 1.800058272,
NA, NA), .Dim=c(8,2) ),
om=structure(.Data=c(.001,0,0,.001),.Dim=c(2,2) ) )
```

The results are shown in Table 2, alongside the results of Haas, and the non-Bayesian multi-variate approach. Note that only four lines of code are required: one to specify the likelihood, two to set the minimally informative priors, and a fourth to monitor TEq. Two additional lines of code invert the precision matrix to obtain the variance-covariance matrix and then obtain a posterior distribution for the correlation coefficient:

s[1:2,1:2] <- inverse(P[,]) r <- s[1,2]/sqrt(s[1,1]*s[2,2])

The predictive distribution given by the MCMC method is considerably wider than the classical methods as it incorporates both the uncertainty due to sampling from a multi-variate distribution, and the uncertainty in its variance-covariance structure. The ratio of the 90th to the 10th centiles in increased from about 47 to about 99.

Scatter plots of samples drawn from the posterior distribution of parameters (Figure 1) portray the nature of the joint uncertainty, and make the point that, conditional on the available evidence, the distributions of the means, standard deviation and correlation have a distinct structure. Extreme values of the mean are associated with high

standard deviation, and high correlation. This structure has significant implications, as high variance and high correlation both widen the predictive distribution of TEq.

4.2. Multi-variate t distribution

One way of exploring the assumption of multi-variate normality is to fit a distribution with heavier tails. Table 2 shows the results for a t distribution with 10 degrees of freedom, achieved by replacing line 1 with:

for (i in 1:8){ y[i,1:2] ~ dmt(mu[],P[,]),10 } # 1. multivariate t- likelihood

The effect is to widen the predictive intervals still further (Table 2), with the ratio of 90^{th} to the 10^{th} centiles now 140.

4.3 Incomplete data

We now consider a situation where additional data is available from other incinerators, but this is not paired (Table 3). This 'incomplete' data structure has the same form as that examined by Hattis and Burmaster. Note that although the unpaired information does not contribute directly to the estimate of the correlation, it does provide indirect information by contributes to the estimates of the variances. To allow comparison with the Table 1 measurements are used twice in a 28-item dataset from 21 incinerators, so that the means and variances are the same. No alterations in the programme are required for this analysis other than changes in the array index from 8 to 22. As expected, uncertainty in TEq is considerably reduced, largely because of reduced uncertainty in the means. Note also, however, that the posterior distribution of the correlation coefficient is very slightly more narrow, due to the extra information on the variances.

4.4 Unrelated repeated measures from the same plant

A further extension allows one to incorporate a variable number of repeat measurements from the same plant on each of the dioxins. We assume, implausibly, that the measurements of the dioxins are unrelated: for example that within each incinerator the measurements of each congener were taken at different times. Each data point in Table 1 is replaced, in effect, with a (variable) number of independent observations. A reasonable assumption would be that the observations, indexed k, are normally distributed about an incinerator- and dioxin-specific means, with an incinerator- and dioxin-specific variance:

$$y_{ijk} \sim N(\mu_{ij}, \tau_{ij})$$

with the μ_{ij} and the τ_{ij} to be estimated from the data. We might further assume that these precision parameters arise from a common gamma distribution, whose parameters will be estimated from the data, given vague priors. This device reduces the effective number of parameters being estimated, as the N_i-times-N_j precision estimates 'borrow strength' from each other, just as the estimated means 'borrow strength' from each other though the common covariance structure.

To implement this, the following two lines replace line 1 of the original code:

for (i in 1:ni) { for (j in 1:nj) for (k in 1:nk) { y[i,j,k] ~ dnorm(mu.y[i,j],tau[i,j])} } mu.y[i,1:nj] ~ dmnorm(mu[],P[,]) }

and the following is added for the precision parameters:

for (i in 1:ni) { for (j in 1:nj) { tau[i,j] ~ dgamma(a,b) }} a ~ dexp (.01) b ~ dgamma(.01,.01)

This code will accommodate a variable number of up to N_k repeat measurements on each incinerator / dioxin combination. Example code and data lists are available from the author on request.

4.5 Correlated repeated measures

A more realistic scenario would be to have a varying number of measurements at each incinerator, but with N_k paired observations on the dioxins, as if for example they were measured on the same day. This generates a second, within-incinerator correlation structure. The model for the congener / incinerator mean in the previous example is replaced by a model which specifies that measurements on dioxins within the same incinerators are multi-variate normal, with a second 2 by 2 precision matrix. As before, this could be given an uninformative Wishart prior and estimated from the data. Lines 1-3 of the original code are therefore replaced by:

for (i in 1:11) { for (j in 1:6) { y[i,j,1:2] ~ dmnorm(mu.y[i,1:2],Pw[,])} } # 'within' model

Pw[1:2,1:2] ~ dwish(Ww[,],2) for (i in 1:11) { mu.y[i,1:2] ~ dmnorm(mu[],Pb[,]) } for (k in 1:2) { mu[k] ~ dnorm(0,.001)} Pb[1:2,1:2] ~ dwish(Wb[,],2)

'between' model

Congdon's example 8.20 (page 420) has a virtually identical structure.⁽¹⁴⁾

5. DISCUSSION

The analyses described here should be seen not so much as an alternative method for incorporating correlation between parameters in simulation models, but as a method for simultaneously estimating the correlation structure and then incorporating it. One obvious advantage of the method is that, in the spirit of probabilistic uncertainty analysis, the uncertainty in the correlation structure is carried through and incorporated in the output. Alternative approaches, based on copula relationships or on correlation structures which are first estimated or elicited and then induced in MC simulation, have to rely on a deterministic form of sensitivity analyses in which 'extreme' correlation structures are explored as an alternative to the base-case.⁽⁸⁾ However, identifying correlation structures that deliver extreme results may itself require further complex analysis, such as dependency bounds.⁽¹⁸⁾

A second advantage is that a Bayesian posterior distribution based on an appropriate characterisation of the data structure faithfully reflects the uncertainties inherent in the available evidence. In the examples given the variance and covariance terms both contribute to the final output, but are not independent conditional on the data, and the particular structure of the data determines the structure of the posterior uncertainty. It is difficult to see how the restricted pairing technique, or any method based on inducing fixed correlations can be adapted to be sensitive to the data structure in the required way. For example, in situations where there are both within- and between-incinerator correlations, the two will be positively correlated, and for the simulation to be correct, their joint uncertainties must be propagated through the model. The essential difficulty is that, given any set of data, the posterior distributions of variance, covariance and mean dioxin levels are themselves related. Entering them into a MC simulation as if they were independent, while certainly better than assuming that covariances are zero, still fails to capture their overall joint uncertainty correctly.

A third advantage is the extreme simplicity of programming what are, in effect, textbook examples of Bayesian multi-variate analysis in the WinBUGS package, the ease of sampling from the appropriate predictive distribution in MCMC, and the many benefits of using standard statistical methods such as the availability of model checking facilities and graphical outputs.

However, there are limitations. Firstly, the method as described depends on the existence of primary data. In the absence of source data, the copula or NORTA methods are available, although it is not easy to see how they could be adapted to incorporate uncertainty in correlation structure. A second limitation concerns the form of the source data. Hattis and Burmaster, who were the first to pose the problem that lead to the current study, discuss a dataset with 17 dioxins measured in 10 incinerators. Clearly, a covariance matrix for this data can have no more than 10 dimensions, so that a procedure like a principle components analysis is needed to reduce the dimensionality. However, this dataset was incomplete, with only 120 of the 170 cells filled. Any formal method for correctly representing the uncertainty in this dataset, including a Bayesian MCMC analysis, would be difficult to implement. A further serious limitation is the computational demands of MCMC modelling, which may not be a feasible method for decision problems with several hundreds of parameters.

Haas⁽⁷⁾ noted the sensitivity of the TEq distribution to the form of copula relationship assumed, though this seems comparatively slight compared both to the effects of data structure, and to sensitivity to the underlying distribution. Although the present approach is centred on the Pearson correlation, one way to parameterise model uncertainty might be to use a multi-variate t-distribution as in section 4.2 but allow the data to estimate the degrees of freedom. A better alternative might be to fit Box-Cox transformations⁽¹⁹⁾. A Bayesian analysis would generate posterior distributions for the degrees of freedom or the Box-Cox skew parameter, and effectively 'model average' over this distribution, although these analyses cannot be carried out in the current version of WinBUGS.

The issue of parameter correlation is just one aspect of the wider question: how to correctly reflect in simulation models the uncertainties inherent in all the available information. As long as all the information sources relate to uncorrelated, individual parameters, simple MC simulation has been adequate. However, this method is not generally able to incorporate information on complex functions of parameters. Correlation is one of many types of information that has this property. More frequently encountered in risk and decision problems is information on model outputs. Two approaches to incorporating data on model outputs, or other functions of parameters, are the Confidence Profile Method of Eddy and colleagues,^{(20) (21) (22) (23)} and the Bayesian Monte Carlo procedure.^{(24) (25)} Bayesian MCMC, and specifically, WinBUGS, is an attractive alternative computational framework for many problems of this sort, ^{(26) (27)} and offers a unified, coherent, framework for evidence synthesis in risk and decision models.

FIGURE CAPTION

Figure 1. Scatter plots based on 10,000 samples from the joint posterior distribution, given the Table 1 data, of the mean and standard deviation (sd) of log 1,2,3,7,8-CDD and its Pearson correlation (r) with log 2,3,4,7,8-PCDF.

ACKNOWLEDGEMENT

The authors acknowledge the contribution of colleagues at the Focus Fortnight on Multi-parameter Evidence Synthesis, held by the Centre for Bayesian Statistics in Health Economics (CHEBS), Department of Probability and Statistics, University of Sheffield, May 11-22, 2002

REFERENCES

- 1. P Doubilet, CB Begg, MC Weinstein, P Braun, and BJ McNeill. "Probabilistic sensitivity analysis using Monte Carlo simulation: a practical approach". *Medical Decision Making* 1985;**5**:157-177.
- 2. GC Critchfield and KE Willard. "Probabilistic analysis of decision trees using Monte Carlo simulation". *Medical Decision Making* 1986;**6**:85-92.
- 3. KM Thompson, DE Burmaster, and EAC Crouch. "Monte Carlo techniques for quantitative uncertainty analysis in public health risk assessments". *Risk Analysis* 1992;**12**:53-63.
- 4. Helton JC and Davis FJ. "Sampling-Based Methods". In: Saltelli A, Chan K, Scott EM, eds. *Sensitivity Analysis*. Chichester: Wiley, 2000;101-154.
- 5. T Bedford and R Cooke. "*Probabilistic Risk Analysis: Foundations and Methods*". Cambridge, UK: Cambridge University Press, 2001.
- 6. RL Iman and WJ Conover. "A distribution-free approach to inducing rank correlation among input variables". *Communications in statistics: simulation and computation* 1982;**11**:311-334.
- 7. CN Haas. "On modelling correlated random variables in risk assessment". *Risk Analysis* 1999;**19**:1205-1214.
- 8. Clemen RT and Reilly T. "Correlations and copulas for decision and risk analysis". *Management Science* 1999;45:208-224.
- 9. Cario MC and Nelson BL "Modeling and generating random numbers with arbitrary marginal distributons and correlation matrix". 1997.
- 10. Hattis D and D Burmaster. "Assessment of variability and uncertainty distributions for practical risk analysis". *Risk Analysis* 1994;**14**:713-730.
- 11. WR Gilks, S Richardson, and DJ Spiegelhalter. " *Markov Chain Monte Carlo in Practice*". London: Chapman & Hall/CRC, 1996.
- 12. A Gelman, JG Carlin, HS Stern, and DB Rubin. "*Bayesian data analysis*". London: Chapman and Hall, 1995.
- 13. JG Carlin and TA Louis. "*Bayes and empirical Bayes methods for data analysis*". Boca Raton: Chapman & Hall/CRC, 2000.
- 14. P Congdon. "Bayesian Statistical Modelling". Chichester: Wiley, 2001.
- 15. Speigelhalter DJ, Thomas A, Best N, and Lunn D. "*WinBUGS User Manual: Version 1.4*". Cambridge, UK: MRC Biostatistics Unit, 2001.
- AE Gelfand and AFM Smith. "Sampling-based approaches to calculating marginal densities". *Journal Of The American Statistical Association* 1990;85:398-409.

- 17. J Besag. "Spatial interaction and the statistical analysis of lattice systems (with discussion)". *Journal of the Royal Statistical Society (B)* 1974;**36**:192-236.
- 18. Ferson S and Burgman MA. "Correlations, dependency bounds and extinction risks". *Biological Conservation* 1995;**73**:101-105.
- 19. Box GEP and Cox DR. "An analysis of transformations". *Journal of the Royal Statistical Society (B)* 1964;**26**:211-246.
- 20. DM Eddy, V Hasselblad, and R Shachter. "*Meta-analysis by the Confidence Profile Method*". London: Academic Press, 1992.
- 21. DM Eddy, V Hasselblad, and R Shachter. "An introduction to a Bayesian method for meta-analysis: the confidence profile method". *Medical Decision Making* 1990;**10**:15-23.
- 22. DM Eddy, V Hasselblad, and R Shachter. "A Bayesian method for synthesising evidence: the confidence profile method". *International Journal of Technology Assessment in Health Care* 1990;6:31-35.
- 23. DM Eddy. "The confidence profile method: a Bayesian method for assessing health technologies". *Operations Research* 1989;**37**:210-228.
- 24. ME Dakins, JE Toll, MJ Small, and KP Brand. "Risk-based environmental remediation: Bayesian Monte Carlo analysis and the expected value of sample information". *Risk Analysis* 1996; **16**:67-79.
- 25. KP Brand and MJ Small. "Updating uncertainty in an integrated risk asessment: conceptual framework and methods". *Risk Analysis* 1995;**15**:719-731.
- 26. DJ Spiegelhalter, JP Myles, DR Jones, and KR Abrams. "Bayesian methods in Health Technology Assessment". 2000.
- 27. AE Ades and S Cliffe. "Markov Chain Monte Carlo estimation of a multiparameter decision model: consistency of evidence and the accurate assessment of uncertainty". *Medical Decision Making* 2002;**22**:359-371.

Incinerator	1,2,3,7,8- CDD	2,3,4,7,8- PCDF	Ln (1,2,3,7,8- CDD)	Ln (2,3,4,7,8 -PCDF)	
1	6.38	53.9	1.853	3.987	
2	18.8	56.4	2.934	4.033	
3	2.4	3.63	0.876	1.289	
4	0.227	11.9	-1.483	2.477	
5	281	190	5.638	5.247	
6	0.663	5.38	-0.411	1.683	
7	1.66	6.05	0.507	1.800	
Mean			1.416	2.930	
Std Dev			2.352	1.496	
Spearman Rank correlation			0.6786		
Pearson correlation			0.8254 (96% CI : 0.1912 - 0.9734)		

Table 1. Data on solid waste emission factors form Haas (1999) ⁽⁷⁾ from seven solid waste incinerators.

Percentile	Haas (min - max)	MC from 'known' multi- variate normal	MCMC Multivariate-normal	MCMC Multi-variate t, df=10	MCMC 21 incinerators, 7 with paired data
2.5%	0.74 - 0.76	0.60	0.35	0.23	0.56
10.%	1.91 - 2.04	1.65	1.43	1.24	1.74
50%	11.4 - 13.5	12.7	12.8	12.7	12.7
90%	88.8 - 93.8	110	141	174	108
97.5%	263 -288	377	865	1637	430
Mean log 1,2,3,7,8-CDD, (sd)		1.42 (0.89)*	1.42 (0.95)	1.38(1.11)	1.41 (0.55)
Mean log 2,3,4,7,8-PCDF, (sd)	•	2.93 (0.57)*	2.93 (0.60)	2.91(0.64)	2.93 (0.35)
Pearson correlation (95% CI)		0.825 (0.19-0.97)	0.806 (0.43-0.97)	0.800 (0.37-0.97)	.806 (0.45-0.95)

Table 2. Centiles of the posterior predictive distribution of TEq, means and posterior sd, from MCMC models, compared to the results of Haas

* Standard error of mean

Incinerator	1,2,3,7,8-CDD	2,3,4,7,8-PCDF
1	6.38	
2	18.8	
3	2.4	
4	0.227	
5	281	
6	0.663	
7	1.66	
8	6.38	53.9
9	18.8	56.4
10	2.4	3.63
11	0.227	11.9
12	281	190
13	0.663	5.38
14	1.66	6.05
15		53.9
16		56.4
17		3.63
18		11.9
19		190
20		5.38
21		6.05

Table 3. Incomplete data structure: 21 Incinerators: 7 with information on 1,2,3,7,8-CDD, 7 with information on 2,3,4,7,8-PCDF, and 7 with information on both.

