COMPARISON OF SAMPLING TECHNIQUES ON THE PERFORMANCE OF MONTE-CARLO BASED SENSITIVITY ANALYSIS

Iain A Macdonald

National Research Council Canada
Institute for Research in Construction
Ottawa, Canada
irc.nrc-cnrc.gc.ca
Contact: iain.macdonald@nrc-cnrc.gc.ca

ABSTRACT

Sensitivity analysis is a key part of a comprehensive energy simulation study. Monte-Carlo techniques have been successfully applied to many simulation tools. Several sampling techniques have been proposed in the literature; however to date there has been no comparison of their performance for typical building simulation applications.

This paper examines the performance of simple random, stratified and Latin Hypercube sampling when applied to a typical building simulation problem. An integrated natural ventilation problem was selected as it has an inexpensive calculation time thus allowing multiple sensitivity analyses to be undertaken, while being realistic as wind and temperature effects are both modeled.

The research shows that compared to simple random sampling: LHS and stratified sampling produce results that are not significantly different (at a 5% level) with increased robustness (less variance in the mean prediction). However, it should not be inferred from this that fewer simulation runs are required for LHS and stratified sampling. Given the results presented here and in previous work it would indicate that for practical purposes Monte-Carlo uncertainty analysis in typical building simulation applications should use about 100 runs and simple random sampling.

INTRODUCTION

The field of sensitivity analysis is becoming more commonplace in building simulation. Early work by Lomas and Eppel (1992) compared the performance of three techniques (differential, Monte-Carlo and stochastic sensitivity analysis). Following on from this Macdonald (2002) embedded sensitivity analysis in ESP-r and de Wit (2001) applied the Monte-Carlo technique to the analysis of natural ventilation and thermal comfort. Since then several authors have used the Monte-Carlo technique on a diverse range of building simulation applications (for example: Hyun et al 2007 and Kotek et al 2007).

To date the effect of sampling technique on the results has not been analysed in the above publications. It is known that stratified sampling can introduce an unknown bias into the results of the analysis (discussed briefly by Macdonald (2002) and de Wit (2001)) and that there can be varying degrees of success with Latin Hypercube (Saltelli et al 2000). This paper will examine three sampling techniques (simple random, stratified and Latin hypercube) and compare their performance.

Sampling techniques

It is standard statistical procedure to use sampling techniques to improve the coverage of the sample, especially when the function being analysed is expensive. The aim of a sampling strategy is to reduce the variance in the estimate of the mean. When applied to building simulation this is attractive as the evaluation of the simulation results (e.g. building energy consumption) can be costly, therefore any reduction in the number of simulations required for a Monte-Carlo analysis will result in a reduction in computational effort.

Three sampling techniques will be described: simple random sampling. Stratified sampling and Latin Hypercube sampling (LHS).

Simple random sampling

This is the most basic sampling technique described here and will be used as the basis for comparisons. The method works by generating a random number and scaling this to the target variable via its probability distribution. The method conforms to the laws of statistics. The mean of the estimate is:

$$\bar{a}_{simple} = \frac{1}{N}\sum_{i=1}^{N} a_i \qquad (1)$$

where there are N samples and a_i is the model output. The sample variance is:

$$Var(a_{simple}) = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \bar{a}_{simple})^2 \qquad (2)$$

and is an unbiased estimate of the population variance. Finally, the variance in the estimate of the mean is:

$$Var(\bar{a}_{simple}) = \frac{1}{N} Var(a_{simple}) \qquad (3)$$

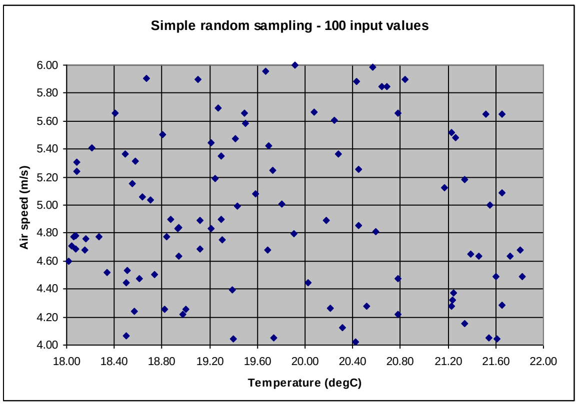

Figure 1. Simple random sampling – 100 input values.

Stratified sampling

This method represents an improvement over simple random sampling by forcing the sample to conform to the whole distribution being analysed. To achieve this, the probability distribution of the target variable is typically divided into several strata of equal probability, one value is then chosen at random within each stratum. From Helton and Davies (2003) the mean of the estimate is:

$$\bar{a}_{stratified} = \sum_{i=1}^{L} \omega_i a_i = \frac{1}{N}\sum_{i=1}^{N} a_i \qquad (4)$$

where there are L strata each with one output (i.e. $N=L$ samples in total), ω_i is the fractional weight of the total population covered by stratum L and a_i are model outputs. The estimate of the variance is:

$$Var(a_{stratified}) = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \bar{a}_{stratified})^2 \qquad (5)$$

Finally, the variance in the estimate of the mean is:

$$Var(\bar{a}_{stratified}) = Var(\bar{a}_{simple}) - \frac{1}{N^2}\sum_{i=1}^{N}(a_i - \bar{a}_{stratified})^2 \qquad (6)$$

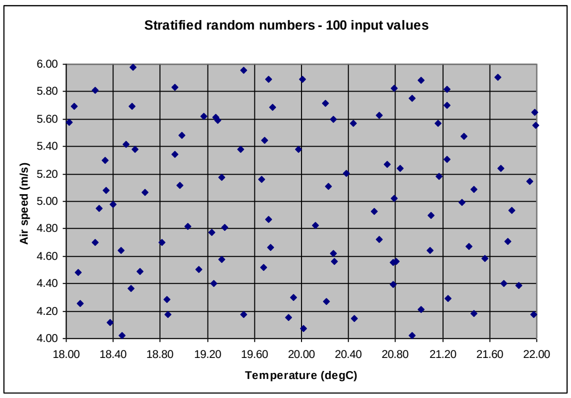

Figure 2. Stratified sampling – 100 input values.

Latin Hypercube sampling

This method is an evolution of stratified sampling that can outperform simple random sampling (Salteilli et al 2000). The method works by dividing the input into strata and then generating samples so that the value generated for each parameter comes from a different stratum (Helton and Davies 2003). For a monotonic function the method has been shown to be better than simple random and stratified sampling (McKay et al 1979), however for non-monotonic functions a bias of unknown size is introduced into the results. The mean and variance of the estimate are the same as for simple random sampling (Saltelli et al 2000, Helton and Davies 2003):

$$\bar{a}_{LHS} = \frac{1}{N}\sum_{i=1}^{N} a_i \qquad (7)$$

$$Var(a_{LHS}) = \frac{1}{N-1}\sum_{i=1}^{N}(a_i - \bar{a}_{LHS})^2 \qquad (8)$$

However, the variance in the estimate of the mean is (Iman 1999):

$$Var(\bar{a}_{LHS}) = \frac{1}{N}Var(a_{LHS}) + \frac{N-1}{N}Cov(A_1, A_2) \qquad (9)$$

where term Cov() is the covariance between the input variables. Thus for LHS to reduce the variance in the mean the covariance should be negative. Iman (1999) quotes results showing that as the sample size tends to infinity this term is nonpositive.

Comparison of techniques

Equations 1 through 9 refer only to the results of each simulation and the number of simulations, thus it can be concluded that the robustness of the Monte-Carlo method is independent of the number of input parameters (this is discussed elsewhere, for example Lomas and Eppel 1992 and Macdonald 2002).

In the context of this paper the robustness of the method can be measured via the variance in the mean output statistic (equations 3, 6 and 9). This variance was estimated from repeated Monte-Carlo analyses as will be described later.

Employing stratification (stratified sampling or LHS) forces the input variable coverage to be better than for simple random sampling. For instance compare Figure 1 and Figure 2; the gridlines represent the 10 strata for each of the two variables being sampled. Figure 1 has several strata without a sample (e.g. the top right corner) and several with more than one sample (e.g. lower right corner). This difference becomes less significant with a larger sample, Figure 3 (Iman 1999). It is expected that the curve for stratified sampling falls between the two curves in Figure 3. The difference after 100 simulations is 2%. Therefore, it should be expected that the results should converge as the sample size increases in terms of the mean and variance in the output.

Several sources, for example Maxval (2009), state that LHS is far superior to simple random sampling. However, Salteilli et al (2000) state that it is not possible to make firm conclusions. Additionally, Hyun et al 2007 state that the minimum number of simulations required is a function of the number of uncertain parameters – this contradicts the theory presented above and appears to come from a discussion of correlation control by Wyss and Jorgensen (1998). A systematic test of sampling techniques is therefore justified in the context of building simulation.

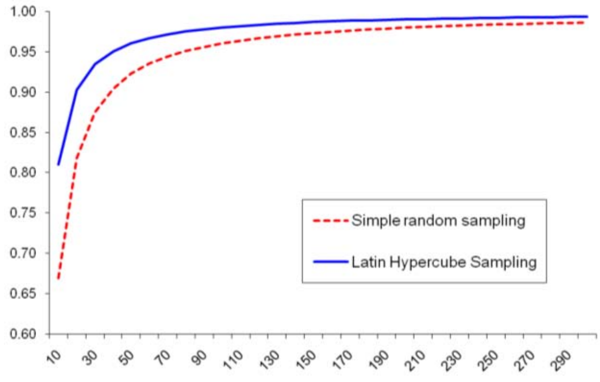

Figure 3. Sample coverage for simple random sampling and LHS.

HYPOTHESES

Three hypotheses will be tested:

1. LHS is more robust than stratified sampling which in turn is more robust than simple random sampling. Robustness is defined as a significantly reduced variance between repeated analyses (i.e. Var($\bar{a}$) will be compared).
2. The bias introduced by stratified and LHS will be insignificant (using simple random sampling as the reference method).
3. The mean and variance will not be significantly different between the methods after 100 simulations.

TEST MODEL

Several algorithms exist to calculate infiltration rates from basic qualitative and quantitative descriptive data for houses (ASHRAE 2005). Recent work (Reardon 2007) compared the availability of data for Canadian houses and that required for available models. It concluded that the single-cell Shaw model (Shaw 1987) was suitable for most analyses. For the purpose of this study the model is quick to run and suitably complex to be representative of a range of building simulation problems.

The essence of the Shaw model is that the infiltration rates due to stack pressure and wind effect are calculated individually and then combined according to superposition of their driving pressures to produce a total infiltration rate. The key input parameters are: wind speed, ambient temperature, internal temperature, neutral pressure ratio and data from a blower door test (coefficients C and n from the curve fit). The model is now elaborated.

Stack-effect driven infiltration is modeled by the equation below:

$$I_S = 0.5\left(\frac{C}{V}\right)\left(\frac{h}{H}\right)(T_{in} - T_{out})^n \qquad (10)$$

Where:

0.5 factor has the units $[m^3 \bullet s \bullet Pa^n)/(L \bullet hr \bullet K^n)]$,

I_S = infiltration air change rate due to stack effect [ac/hr],

C = house flow coefficient from curve fit of the leakage test data $[L/(s \bullet Pa^n)]$,

V = internal volume of the house including basement $[m^3]$,

h = height above grade of the neutral pressure level [m],

H = height above grade of the upper ceiling of the house [m],

T_{in} = indoor air absolute temperature [K],

T_{out} = outdoor air absolute temperature [K], and

n = house flow exponent from curve fit of the leakage test data.

Shaw suggests that for a house without a flue h/H = 0.64, and for a house with a single 127 mm dia. flue h/H = 0.86, based on the data set used to develop this model. A later study (Reardon 1989) with measured NPLs in a larger number of houses has provided a guide for NPL=0.6 for houses without an open flue and 0.7 with an open flue.

The form of the curve fit to the leakage test data (from a fan depressurization measurement of the envelope airtightness of the house, following CGSB 1986), that is used to determine the flow coefficient and flow exponent is the power law curve:

$$Q_m = C(\Delta P)^n \qquad (11)$$

Where:

Q_m = measured flow rates [L/s], and

ΔP = measured pressure difference across envelope [Pa].

Wind driven infiltration is modeled by the equation below:

$$I_W = 0.4\left(\frac{C}{V}\right)U'^{2n} \quad \text{Exposed} \qquad (12)$$

$$I_W = 0.7\left(\frac{C}{V}\right)U'^{n} \quad \text{Shielded} \qquad (13)$$

Where:

0.4 factor has units $[(m^3 \bullet Pa^n \bullet s^{2n+1})/(L \bullet hr \bullet m^{2n})]$,

0.7 factor has units $[m^3 \bullet Pa^n \bullet s^{n+1})/(L \bullet hr \bullet m^n)]$,

I_W = infiltration air change rate due to wind [ac/hr], and

U' = windspeed measured at height of 20m on-site [m/s].

The combined infiltration due to both stack-effect and wind is modeled by combining these two component infiltration rates using n-quadrature to effectively superpose the pressures created by these two physical phenomena, since the two component infiltration rates do not simply add, due to the non-linear relationship between driving pressure and driven flow rate. The combined model equation is:

$$I_{WS} = F\left(I_S^{(1/n)} + I_W^{(1/n)}\right)^n \qquad (14)$$

Where:

I_{WS} = total combined infiltration air change rate [ac/hr],

F = an empirical factor defined by the following:

$$F = 1 \text{ for } 0 \le \frac{I_{sml}}{I_{lrg}} < 0.1 \qquad (15)$$

$$F = 0.8\left(\frac{I_{sml}}{I_{lrg}}\right) - 0.1 \text{ for } 0.1 \le \frac{I_{sml}}{I_{lrg}} < 1 \qquad (16)$$

Where:

I_{sml} = the smaller of the two components I_S and I_W, and

I_{lrg} = the larger of the two components I_S and I_W.

METHOD

Data from a single real building in Quebec City, Canada was used as the input data for the test model. The sensitivity analysis was undertaken for a range of ambient temperatures (18-22 m/s) and wind speeds (4-6 m/s) using the simple random, stratified and latin hypercube sampling strategies. Figure 4 shows a typical cumulative distribution for the input variables for the simple and stratified sampling compared to the 'ideal'. The results are from a 100 point sample and totalled in 10 bins. As expected the stratified sampling technique produces a better distribution compared to the simple random sample (i.e. the line is straighter).

For all cases the Monte-Carlo analysis used a sample of N_{runs} simulations to generate the output distribution. This analysis was then repeated $M_{repetitions}$ times and the range of results calculated.

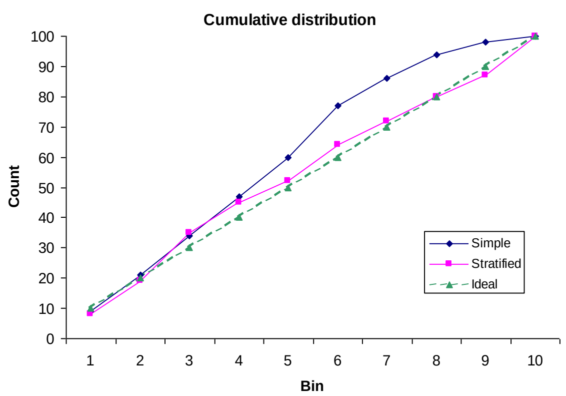

Figure 4. Typical cumulative distribution for the input variables.

To examine the differences in predictions resulting from the different sampling procedures the process depicted in Figure 14 was adopted. For each sampling option the mean and variance in the predicted infiltration rate were predicted. Four sampling options were used:

1. Simple random sampling. To guarantee a suitable coverage 100 samples were generated, thus N_{runs}=100.
2. Stratified sampling. The input distributions for both variables were divided into 10 equal probability strata. A sample was drawn at random from each stratum, to generate 100 input samples, thus N_{runs}=100.
3. Latin Hypercube Sampling (LHS(10)). Using the same strata as in case 2 ten input samples were generated, thus N_{runs}=10.
4. Replicated Latin Hypercube Sampling (LHS(10x10)). The procedure for LHS(10) was repeated 10 times, each time a new random pairing was generated for the two input variables, thus N_{runs}=100.

This process was initially repeated 10 times (i.e. $M_{repetitions}$=10) for each of the four cases and then 100 times to compare the robustness of the methods.

In addition the simple random sampling method was repeated 100 times for 100, 1000, 10000 and 50000 samples (i.e. $M_{repetitions}$=100 for N_{runs}=100, 1000, 10000 and 50000). These results were used as a benchmark against which the remaining cases could be compared.

RESULTS

Summary statistics were generated from the simulations. For each of the four sampling scenarios the mean result and its variance was calculated (from the $M_{repetitions}$ analyses), in addition the mean of the variance was calculated and its variance. With these statistics the hypotheses can be tested.

Simple Random Sampling

Figure 5 shows the variance in the mean of the four analyses using different sample sizes for simple random sampling only. As can be seen the variance is reduced by increased sampling (i.e. increasing N_{runs}). The significance of this reduction was tested against an F-distribution (Kreyszig 1993). In all three differences the reduction is significant at the 5% level. Therefore, the increased sampling produces a significantly more robust result.

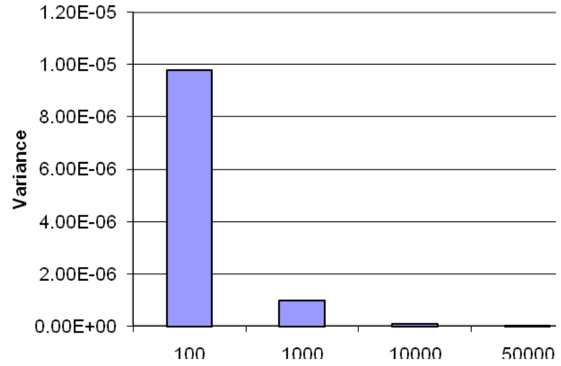

Figure 5. Variance in mean results by N_{runs}.

Examining the mean and variance in the results (Figure 6 and Figure 7 respectively) it can be seen that there is less variation due to the additional runs. Again the significance of these differences was tested at the 5% level (the variance was compared as above and the means against a t-distribution). In all cases the difference was not significant.

Overall, this would indicate that there is little point in running more than 100 simulations if using simple random sampling as the estimates of mean and variance will not vary significantly up to 50000 simulations.

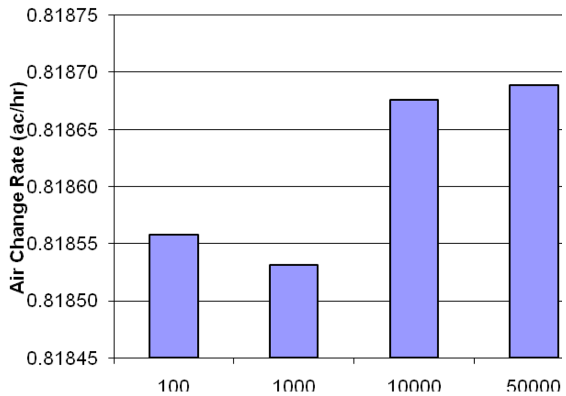

Figure 6. Mean air change rate by N_{runs}.

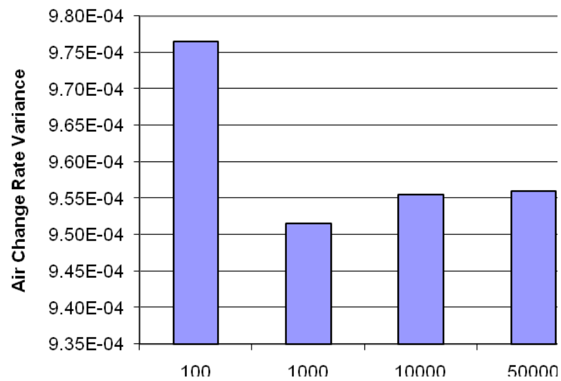

Figure 7. Mean variance in air change rate by N_{runs}.

100 Samples

Comparing the three sampling methods using 100 samples/strata as described above:

1. Figure 8 shows the results for mean air change rate. Focussing on the first three columns there is no significant difference at the 5% level between the results in the first three columns. Therefore, there is no significant bias introduced.
2. Figure 9 shows the mean variance in the results. Again focussing on the first three columns only there is no significant difference at the 5% level between the results in the first three columns.
3. Finally, Figure 10 shows the variance in the mean air change rate (robustness). Again focussing on the first three columns only there is a significant difference at the 5% level between the simple and stratified results and between the simple and LHS(10x10) results, i.e. the use of stratification generates a more robust result compared to simple random sampling.

The results in the final columns of Figure 8, Figure 9 and Figure 10 show that by only using 10 samples as opposed to 10 replications of the LHS method there is no significant difference at the 5% level for the mean and variance results, however the robustness is significantly improved by replication (Figure 10).

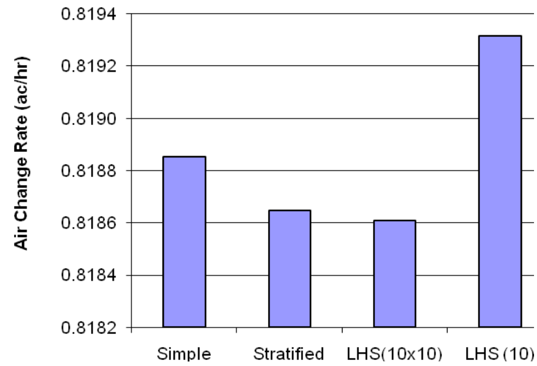

Figure 8. Mean air change rate.

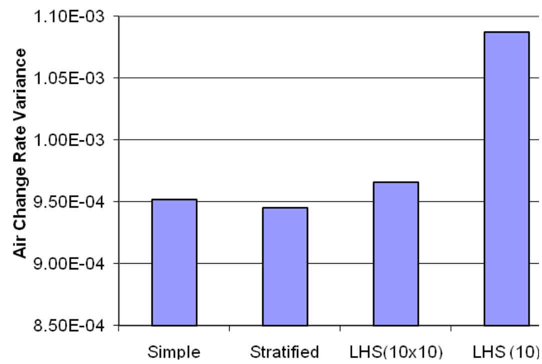

Figure 9. Mean variance in air change rate.

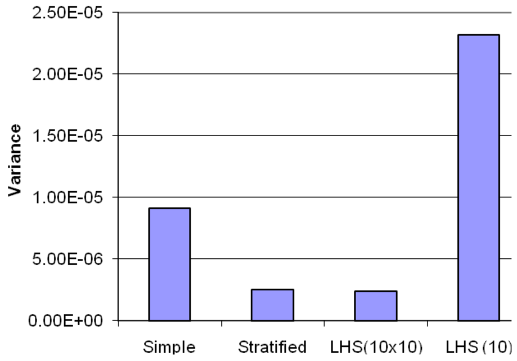

Figure 10. Variance in mean air change rate.

10 Samples

Comparing the three sampling methods using 10 samples for the simple and LHS methods and 9 for stratified (in this case three strata were used for each of the two variables resulting in 9 runs):

1. Figure 11 shows the results for mean air change rate. There is no significant difference between the results at the 5% level.
2. Figure 12 shows the mean variance in the results. Again there is no significant difference in these results at the 5% level.
3. Finally, Figure 13 shows the variance in the mean air change rate (robustness). There is a significant difference at the 5% level between all methods, i.e. the use of stratification generates a more robust result compared to simple random sampling.

Sample size and method comparison

Comparing the results from the 100 run simple sampling case with the 10 run LHS case, at the 5% level, there is no significant difference in the estimation of the mean and variance. However, there is a significant difference in the variance of the mean results (robustness). In this case the larger samples (simple) is more robust than the LHS method. The same results are evident when comparing the stratified and LHS methods.

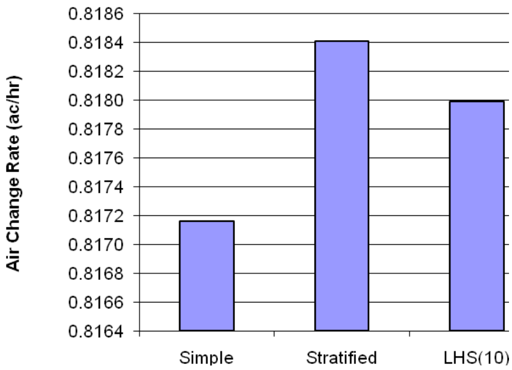

Figure 11. Mean air change rate.

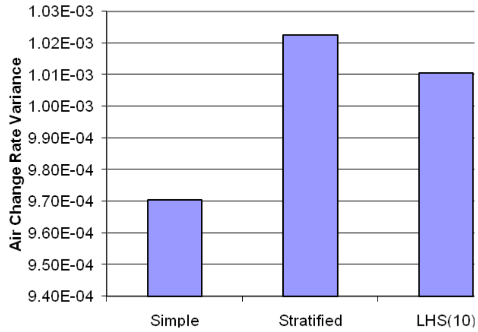

Figure 12. Mean variance in air change rate.

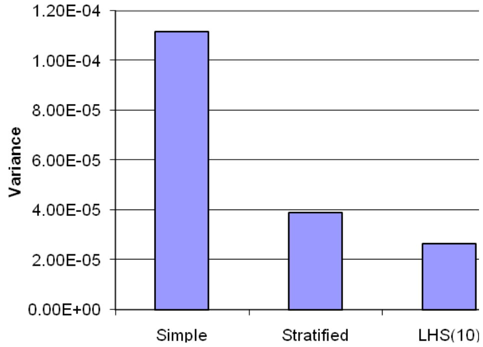

Figure 13. Variance in mean air change rate.

DISCUSSION

This paper set out to test three hypotheses related to the sampling method used for a Monte-Carlo analysis:

Method robustness

The results presented have shown that for equal sample sizes the robustness of the sampling methods was as expected. This was true for the following cases:

- 100 simple < 100 stratified < 10x10 LHS
- 10 simple < 9 stratified < 10 LHS

where the number is the number of samples (and 10x10 represents the replicated LHS runs).

There is also a significant change in the robustness between:

- 10 LHS < 100 simple
- 10 LHS < 100 stratified

and

- 100 simple < 1,000 simple < 10,000 simple < 50,000 simple

Examining these results as a whole would indicate that the non-simple sampling techniques should be used to produce a more robust analysis, but not at the expense of fewer simulation runs.

Bias

In all cases there was no significant difference between the mean results for equal numbers of simulations. This was also true in the comparison between:

- 100 simple and 10 LHS
- 100 stratified and 10 LHS

This would indicate that there is no significant bias introduced by the sampling methods.

Mean and variance

The results presented were not significantly different for the following pairs:

- 100 simple & 1,000 simple
- 1000 simple & 10,000 simple
- 10,000 simple & 50,000 simple
- 100 simple & 100 stratified
- 100 simple & 10x10 LHS
- 100 stratified & 10x10 LHS
- 10x10 LHS & 10 LHS
- 10 LHS & 9 stratified
- 10 simple & 10 LHS
- 10 simple & 9 stratified

This would indicate that with very small samples (9 or 10 simulations) the results from a Monte-Carlo analysis are accurate in terms of the estimate of the mean and variance of the output.

Practical concerns

Given these results it would indicate that for practical purposes there is no requirement for LHS or stratified sampling in typical building simulation problems.

The results presented here are for samples drawn from strata with equal probabilities – for multi-variate problems with multiple probability distributions management of the strata weights could become complex (formulae exist for accounting for the weights but this adds to the complexity of the analysis). Given the simplicity of simple random sampling and that after 100 simulations the coverage of the input range is within 2% of that with LHS, for ease of implementation, the simple sampling method would be preferred.

CONCLUSIONS

This paper set out to examine three aspects of the sampling method used in Monte-Carlo uncertainty analysis. It has shown that:

1. There is no significant bias introduced by the stratified and LHS methods compared to simple sampling.
2. For the analysed cases there is no significant difference in prediction of mean and variance between the methods in terms of sampling method and number of simulations (N_{runs}).
3. For the same number of simulations the LHS method produces a more robust result compared to the stratified method, which in turn produces a more robust result compared to the simple method.
4. The stratified and LHS methods cannot necessarily be used as a mechanism to save simulations runs.

Given these results and the analysis presented by Lomas and Eppel (1992) it would indicate that for practical purposes Monte-Carlo uncertainty analysis in typical building simulation applications should use about 100 runs and simple random sampling.

REFERENCES

ASHRAE. 2005. Handbook of Fundamentals: Chapter 27 Ventilation and Infiltration, Atlanta: ASHRAE.

Helton J.C. and Davis F.J. 2003. “Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems”, Reliability Engineering and System Safety (81), pages 23-69, Elsevier.

Hyun S, Park C, Augenbroe G. 2007. “Uncertainty and Sensitivity Analysis of Natural Ventilation in High-rise Apartment Buildings”, Proc Building Simulation 2007, Beijing, China.

Iman R.L. 1999. “Appendix A: Latin Hypercube Sampling” Encyclopedia of Statistical Sciences, Update Volume 3, Wiley NY.

Kotek P, Jordán F, Kabele K, Hensen J. 2007. “Technique of Uncertainty Analysis for Sustainable Building Energy Systems Performance Calculations”, Proc Building Simulation 2007, Beijing, China.

Kreyszig E. 1993. “Advanved Engineering Mathematics”, John Wiley and Sons, 7th Edition.

Lomas K J, Eppel H. 1992. “Sensitivity analysis techniques for building thermal simulation programs”, Energy and Buildings, no 19.

Macdonald I A, 2002, “Quantifying the effects of uncertainty in building simulation”, PhD thesis, ESRU, University of Strathclyde.

McKay M.D., Beckman R.J. and Conover W.J. 1979. “A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code”, Technometrics (Vol 21, No 2), pages 239-245, American Society for Quality Control and The American Statistical Association.

Madras N. 2002. “Lectures on Monte Carlo Methods”, Fields Institute Monographs, American Mathematical Society.

Maxvalue. 2009. www.maxvalue.com/tip025.htm

Reardon, J.T. 2007. "Assessment of natural ventilation for Canadian residential buildings" CMHC Research Report, CHIC Library Cat.No. Ca1 MH110 07A75, Canada Mortgage and Housing Corporation, Ottawa. (ftp://ftp.cmhc-schl.gc.ca/chic-ccdh/Research_Reports-Rapports_de_recherche/eng_unilingual/CHIC_Assessment_Natural(w).pdf)

Saltelli A, Chan K, Scott E M, “Sensitivity analysis”, J Wiley and Sons, 2000.

Shaw, C.Y. (1987), ‘Methods for estimating air change rates and sizing mechanical ventilation systems for houses', ASHRAE Transactions, 93 (2), pp. 2284-2302, (NRCC-30422) (IRC-P-1597)

de Wit S, 2001. "Uncertainty in predictions of thermal comfort in buildings", PhD thesis, Delft University.

Wyss, G.D. and K. H. Jorgensen. 1998. A User's Guide to LHS: Sndia's Latin Hypercube Sampling software, Albuquerque, NM, Sandia National Laboratories.

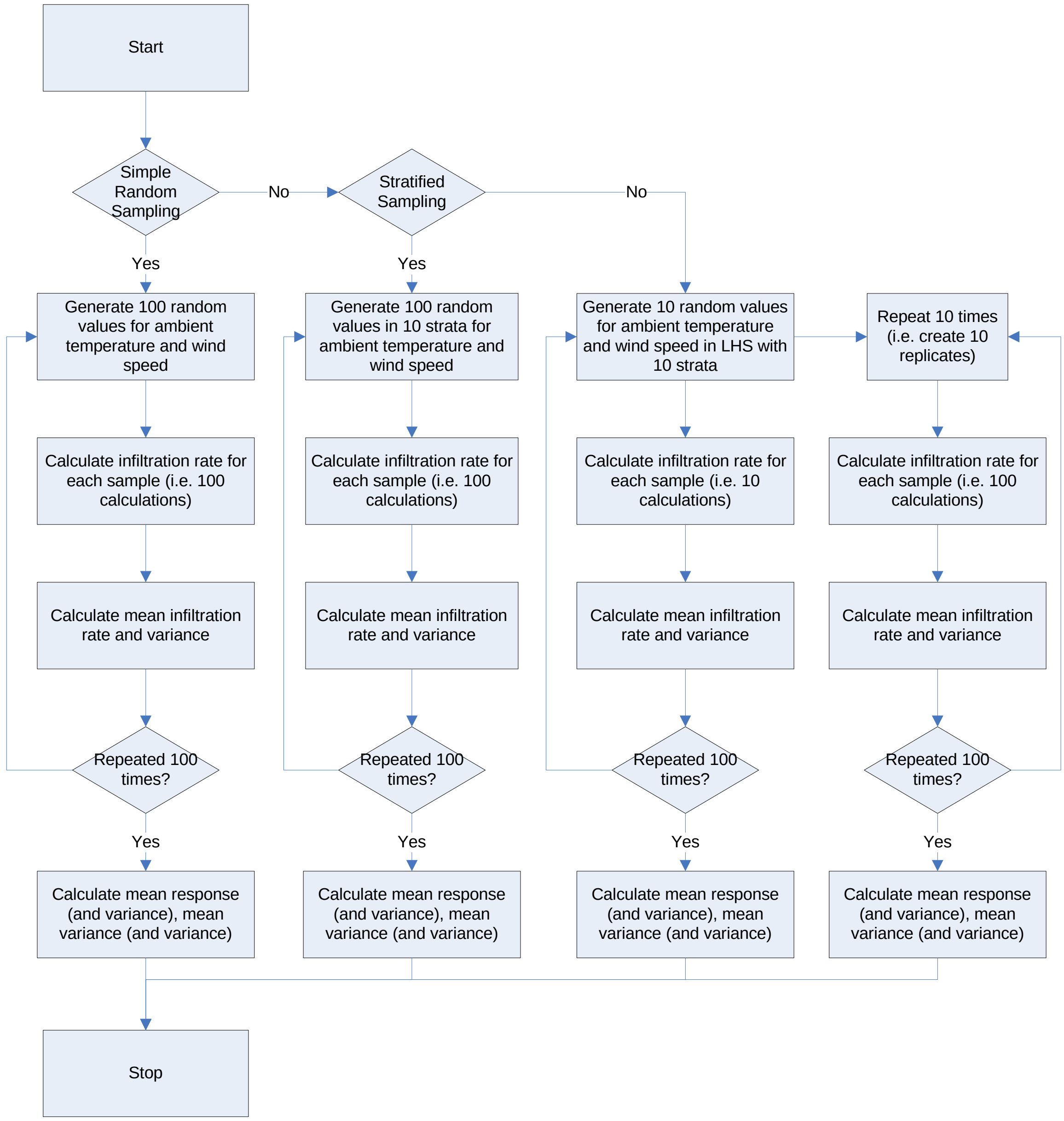

Figure 14. Analysis process for 100 sample runs.