Maximum Simulated Likelihood Estimation with Spatially Correlated Observations: A Comparison of Simulation Techniques

Xiaokun Wang
Graduate Student Researcher
The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
wangxk@mail.utexas.edu

Kara M. Kockelman
(Corresponding Author)
Associate Professor & William J. Murray Jr. Fellow Department of Civil, Architectural and Environmental Engineering The University of Texas at Austin
6.9 ECJ, Austin, TX 78712-1076
kkockelm@mail.utexas.edu
Phone: 512-471-0210
FAX: 512-475-8744

The following paper is a pre-print and the final publication can be found as Chapter 8 in *Transportation Statistics,* Brian Sloboda (Ed.) 2009: 173-189, JD Ross Publications

ABSTRACT

Econometric models are a powerful tool for analyzing regional issues. Complex models are normally intractable and require special estimation methods. Maximum simulated likelihood estimation (MSLE) techniques have become popular in recent years, and are being included in new software releases (such as STATA and Limdep). It is important that analysts understand the relative performance of different simulation techniques under various data circumstances. This especially true in regional studies, where observations are often spatially correlated.

This paper studies the performance of several simulation techniques with spatially correlated observations. Quasi Monte-Carlo (QMC) methods are found to impose a strong periodic correlation pattern across observations. While some forms of sequencing, such as scrambled Halton, Sobol and Faure, can sever correlations across dimensions of error-term integration, they cannot remove the correlation that exists across observations. When a data set's true correlation patterns clearly differ from the simulated patterns, model estimation may become inefficient; and, with finite samples, statistical identification of parameters may suffer. Fortunately, here we find that, at least within the mixed logit framework, even when observations are correlated, QMCs and hybrid methods are typically preferred to pseudo Monte-Carlo methods, thanks to their better coverage. These findings offer an important supplement to existing studies of spatial model estimation and should prove valuable for future work that requires simulated likelihoods with spatially correlated observations.

KEY WORDS: Simulation techniques, mixed logit model, spatial correlation, correlated observations, Halton sequence, scrambled Halton, shuffled Halton.

INTRODUCTION

Simulation is an indispensable tool in many research fields for studies of complex systems. Thanks to enhanced computational power, simulation has become more popular in recent years. One important application is the estimation of statistical models where analytical derivation is impractical (for example, when a likelihood function involves a multidimensional integral). In regional studies, a typical application is the estimation of complex spatial econometric models, especially those that are nonlinear in nature. In this study, simulation techniques are explored within the framework of a mixed (random-coefficients) logit model, where observations are spatially correlated.

McFadden (1989) first applied MSLE for discrete choice. Ideally, simulated draws should be randomly chosen points from the given (typically standard uniform) distribution. In practice, however, random numbers have to be generated using a deterministic routine. For example, the widely used pseudo Monte-Carlo (PMC) method generates numbers in a deterministic fashion, but in a way that appears to be random under simple statistical tests (Niederreiter, 1992). This type of simulation causes integration errors[1] to converge on zero at a rate of $\sqrt{N}$, where N is the number of observations (Spanier and Maize, 1994).

In contrast to PMC, quasi Monte-Carlo (QMC) methods do not generate apparently random values. Their major advantage is that they are uniformly distributed with better coverage, offering a faster convergence rate for the upper bound of integration error (Bhat, 2003). There are many ways to generate QMCs: commonly used approaches are the Faure sequence (Faure, 1992), Sobol sequence (Sobol, 1967), Halton sequence (Halton, 1960) and the Halton's various scrambled and shuffled versions (see, e.g., Braaten and Weller, 1979, Hellekalek, 1984 and Hess and Polak, 2003). As Bhat (2003) explains, a scrambled Halton works better than a standard Halton for high dimension integration. One limitation of QMC is that "there is no practical way of statistically estimating the integration error" (Bhat, 2003, pg 839). Another potential issue, as this paper will discuss, is the periodic correlation pattern inherent in QMC. For samples with independent observations, the effect of this correlation pattern is not significant; however, it may become problematic when the model involves correlation across observations.

As Bhat (2003) suggests, an attractive alternative to PMC and QMC is their combination, also called randomized QMC. Bhat (2003) proposed shifting a Halton sequence by adding a random number to the standard Halton sequence. In this way, the integration error can be statistically estimated. However, shifting will not break the cycling of a standard Halton sequence. Thus, it does not help remove the correlation pattern (across observations) in QMC. Hess and Polak (2003) suggested another hybrid method, a shuffled Halton sequence. By randomly shuffling (reordering) a short Halton sequence for each observation, sequences avoid correlation in higher dimensions while speeding sequence generation time. However, because the shuffled Halton sequence uses the same numbers (though in different orders) for all observations, the simulation noise is likely to increase with sample size, instead of canceling. As discussed here, this method may lead to inconsistent estimates.

In summary, many studies seek improved simulation techniques to produce unbiased, consistent, and maximally efficient estimates. Most existing evaluations of simulation techniques focus on asymptotic properties using assumptions of inter-observational independence (see, e.g., Ben-Akiva et al., 2001, Bhat, 2003 and Hess and Polak, 2003). When there are repetitive choice experiments, researchers tend to add an individual effect into the model specification (see, e.g., Hensher and Green, 2003).In regional studies, observations are often spatially correlated. For example, in models of land use change, land development of a

[1] Integration error, or discrepancy, is the difference between true value (derived via integration) and the simulation-derived approximation.

specific site often depends on neighborhood conditions. In other data sets (e.g., traffic counts), observations are interrelated through network connectivity. Spatial and serial autocorrelation in such situations are systematic. In these cases, simulation techniques relying on an independence assumption may be problematic; simply specifying unit-specific random effects is not adequate either.

Thus, the objective of this study is to explore various simulation techniques in circumstances where continuous and systematic correlation across observations exists. Six simulation methods are compared: PMC, standard Halton, scrambled Halton, shuffled Halton and two other hybrid methods referred to here as "long shuffled Halton" method and "randomized shuffled Halton" sequence. (Two other QMC sequences, Sobol and Faure, are also briefly discussed, but not further compared because of their similarity to standard Halton sequences.)

As explained in greater detail below, the "long shuffled Halton" sequence has a generation rule similar to the shuffled Halton. Instead of shuffling the same sequence for all observations, the "long shuffled Halton" approach shuffles a long standard Halton sequence and then splits it for each observation. The "randomized shuffled Halton" is a hybrid version of Bhat's (2003) shifted Halton sequence and Hess and Polak's (2003) shuffled Halton sequence: It adds a uniformly distributed random number to each value in a shuffled sequence. The relative performance of these six techniques is then compared based on theory and an empirical study using synthetic data. As this paper highlights, standard and scrambled Halton methods import periodic correlation across observations. Thus, in situations where correlated observations exist, Halton and scrambled Halton sequences may cloud or counteract the true correlation pattern, leading to inconsistent (or at least inefficient) estimates.

The following sections describe existing studies of mixed logit model estimation and simulation techniques, along with potential problems for different simulation methods. These methods are applied to a synthetic dataset exhibiting spatial correlation. The results of different simulation methods are compared.

MIXED LOGIT MODEL

The mixed logit model, also called the random parameters or kernel logit model, has been widely used for a number of years. It has been applied to topics in finance, biometrics and social science, among others. Its application in regional science is also very broad. For example, Rouwendal and Meijer (2001) used it to study preferences for housing and jobs, Wang and Kockelman (2006) used it to study the spatial and temporal evolution of land cover in urban area, and Hensher and Greene used it to analyze urban commuting.

Initially, the mixed logit model was designed to incorporate heterogeneity and correlation across alternatives. Thanks to its great flexibility, its applications were extended. Srinivasan and Mahmassani (2005) proved that mixed logit models are capable of approximating any random utility model. And Greene (2002) notes how mixed logit models can conveniently incorporate panel effects.

McFadden and Train's (2000) work depicts a general formulation for mixed logit models, where parameters are assumed to be randomly distributed with a density distribution function $f(\beta \mid \theta)$. β denotes the random coefficients, and θ parameters characterize their random distribution. For observation i, the probability of choosing alternative q is

$$P_{iq} = \int \left(\exp\left(x_{iq}'\beta\right) \Big/ \sum_{p} \exp\left(x_{ip}'\beta\right) \right) f\left(\beta \mid \theta\right) d\theta \qquad (1)$$

where x_{iq} is the vector of explanatory variables. This integration can be approximated using simulation:

$$SP_{iq} = \frac{1}{R}\sum_{r=1}^{R} P_{iq}\left(\beta^{ir}\right) \tag{2}$$

where SP_{iq} is the simulated probability for observation i's choice of alternative q, β^{ir} is the *rth* of *R* draws for observation i from its density $f\left(\beta \mid \theta\right)$. By construction, SP_{iq} is an unbiased estimate of P_{iq}; and, therefore, the simulated likelihood value SL_i is also an unbiased estimator of the true likelihood value L_i: $E\left(SL_i\right) = L_i$.

The log likelihood of observation i choosing alternative q is

$$\ln L_i = \sum_{p=1}^{J} d_{ip} \ln P_{ip} \tag{3}$$

where d_{ip} equals 1 if observation i chooses alternative q, and 0 otherwise. Unfortunately, the log transformation leads to bias in the simulated log likelihood. As Train (2003) shows,

$$\ln SL_i = \ln L_i + \left(E\left(\ln SL_i\right) - \ln L_i\right) + \left(\ln SL_i - E\left(\ln SL_i\right)\right) \tag{4}$$

The second part of this expression $\left(E\left(\ln SL_i\right) - \ln L_i\right)$ indicates the bias caused by the transformation, denoted here as B. A second-order Taylor expansion suggests that

$$B = E\left(\ln SL_i\right) - \ln L_i \approx -\frac{1}{2L_i^{2}} \operatorname{var}\left(SL_i\right) \tag{5}$$

As Train (2003) explains, $\operatorname{var}\left(SL_i\right)$ is inversely proportional to the number of draws, *R*. Therefore, when *R* is fixed, this bias *B* will accumulate with sample size, making MSLE inconsistent. Only with an increase in *R* will bias disappear.

The third part of equation (4), $\ln SL_i - E\left(\ln SL_i\right)$, defines the simulation noise, which is caused by the difference between draws actually used for SL_i and its expectation over all possible draws. It is referred to here as *C*, which also has a variance inversely proportional to *R*. As long as SL_i is an unbiased estimator of L_i, the major concern is to reduce the summation of B and C over the entire sample. With PMC, this can always be achieved by increasing *R*. The remaining issue is to find the simulation technique that achieves these goals most efficiently. Train (2003) suggests that if draws are negatively correlated within one observation and/or across observations, the sample summation of B and C tends to fall. Train (2003) supported this assertion by using R=2 as an example: The variance of $\hat{t} = \left(t\left(\varepsilon_1\right) + t\left(\varepsilon_2\right)\right)/2$ is $\left(\operatorname{var}\left(t\left(\varepsilon_1\right)\right) + \operatorname{var}\left(t\left(\varepsilon_2\right)\right) + 2\operatorname{cov}\left(t\left(\varepsilon_1\right), t\left(\varepsilon_2\right)\right)\right)/4$: A negative $\operatorname{cov}\left(t\left(\varepsilon_1\right), t\left(\varepsilon_2\right)\right)$ reduces the variance of average. Negative correlation within a given observation's draws, which can reduce B, actually provides good coverage. Correlation across observations, which influences the total variance of simulation noise C, depends on the relationship across draws used for different observations. Based on all these relationships, studies of simulation techniques seek a simulation method that provides good coverage and negative correlation across observations.

SIMULATION TECHNIQUES

Standard Halton sequence

As previously mentioned, Halton sequences (Halton,1960) are deterministic series of values between zero and one that provide uniform coverage in this number space. Each sequence is generated via a prime number, also called its base. Bhat (2003) explains the generation of Halton sequences with formulations, and Train (2003) illustrated how a standard Halton sequence draws numbers cyclically on the [0,1] interval. To make the following explanation clearer, Train's (2003) illustration is briefly described here:

Using prime number 3 as an example, the first cycle divides the interval into three sections, their left edges labeled as A, B, and C (shown in Figure 1). Then each section is divided into three parts, and each subsection is labeled lexicographically, as for the first cycle. The Halton sequence is generated with a modified alphabetical order: B for 1/3, C for 2/3, BA for 1/9, BB for 4/9, BC for 7/9, CA for 2/9, CB for 5/9 and CC for 8/9. The draws continue cycling in this way, filling in the interval's gaps. Zero and duplicated points (e.g., AA=A=1, AB=B, AC=C) are ignored.

Train (2003) noted that Halton sequences exhibit a negative correlation within one observation's sequence, as well as across adjacent observations. Thus, Halton sequences can perform much better (in reducing simulation bias and noise) than independent draws. However, one potential issue is the incomplete consideration of cross-observation correlation. When additional correlation terms are considered, the conclusions may change. For example, if R=3, $\operatorname{cov}\left(t\left(\varepsilon_1\right),t\left(\varepsilon_2\right)\right)<0$ and $\operatorname{cov}\left(t\left(\varepsilon_2\right),t\left(\varepsilon_3\right)\right)<0$, it is quite possible that $\operatorname{cov}\left(t\left(\varepsilon_1\right),t\left(\varepsilon_3\right)\right)>0$. If this positive value exceeds the sum of those two negative values, the final variance will be larger than the variance resulting from independent draws. Unfortunately, this is quite common with Halton sequences. As illustrated by Figure 1, a Halton sequence's cyclical draws pick up the spaces left by prior draws and ultimately pick positions close to a series of earlier draws. Thus, it is reasonable to expect positive periodic correlations. Wang and Kockelman (2006) noted this pattern across draws and touched on it in their mixed logit models of land use change, based on satellite imagery. In their study, the correlation across observations is discussed based on the correlation between generated sequences. Here, the correlation across observations' simulated probabilities is examined in depth using a method Train (1999) described: A single observation is treated as 1000 different observations, so that the variation pattern in variables is avoided. In this way, the correlation across observations' simulated probability is exclusively generated by the simulation technique.

The difference between this method and the one described by Wang and Kockelman (2006) is that when calculating correlations between two sequences of random numbers, the order of these numbers is important. For example, consider a shuffled Halton sequence, where the randomly ordered sequences show no correlation across observations. Simulated probabilities, in contrast, diminish the re-ordering effect by averaging a transformation of the simulated values over the set of draws: $SP_{iq}=\frac{1}{R}\sum_{r=1}^{R}P_{iq}\left(\beta^{ir}\right)$. This means that when using simulated probabilities, the order of r has no influence. From this perspective, Train's (1999) method indicates that shuffled Halton sequences result in perfectly correlated simulated probabilities.

In order to clearly illustrate the effect of simulation technique on observation correlation, a simple model specification was used: A three-alternative mixed logit model with two alternative-specific constant (ASC) terms and one alternative-specific variable. The alternative-specific variable was generated from a standard uniform distribution and has a fixed coefficient 1. The first alternative serves as the base or reference alternative, and thus has a fixed ASC of zero. Alternative two's ASC is fixed as 0.5. The ASC for alternative three is normally distributed with mean 0.2 and standard deviation 2. .

Figure 2 shows the correlation pattern generated using different numbers of draws (*R*) from the same base of 3 for alternative 3. The vertical axis shows correlation of simulated probabilities, $corr\left(SP_{\bullet,q}, SP_{\bullet+k,q}\right)$, where $SP_{\bullet,q}$ is the vector of all observations' simulated probabilities for choice of alternative q. (Figure 2 shows the pattern for *q*=3; the *q* = 1 and 2 patterns are virtually the same.) The numbers on the horizontal axis indicate the "distance" between observations, i.e., k. That is, 1 indicates adjacent observations, while 3 means every third observation. Similar to Train's (1999) findings, the correlation between adjacent observations is around -0.4. However, correlation between pairings of every third observation jumps to +0.6. In fact, with *R*=25, the same results emerge for all numbers that are multiples of 3. As the number of draws increases, the variations in correlation values softens a bit, but for certain numbers (e.g., *k*=27), the correlation remains quite high.

The base 3 results are very similar to those seen in Halton sequences generated using other prime numbers (i.e., other bases). As shown in Figure 3, the correlations across observations always present approximately periodic patterns, with multiples of the bases always associated with high positive values, and the peak correlations can be very high. For example, when the base is 5 and *R* equals 100, the correlation between observations and their 25^{th} nearest observations is 0.975. This means that Halton sequences impose a strong correlation pattern on observations' simulated probabilities. Within a given model specification, the cycle and the magnitude of this periodic correlation depend on its base and the number of draws.

When observations are independent, summation can cancel out these negative and positive correlations, making this effect of correlation across observations insignificant. Meanwhile, the good coverage of a Halton sequence within an observation helps to reduce the value of Equation (5). Thus, when observations are independent, the Halton sequence is generally preferred to PMC. Munizaga and Alvarez-Daziano (2001), Bhat (2003), Hess and Polak (2003) all confirm this conclusion.

However, if observations are correlated, this correlation pattern imposed by use of Halton sequences becomes more problematic.

Scrambled Halton Sequence

Bhat (2003) proposed application of scrambled Halton sequences to estimate mixed logit models. There are many ways to scramble a Halton sequence (e.g., Braaten and Weller, 1979 and Hellekalek, 1984). The scrambling method discussed here is the one proposed by Braaten and Weller (1979), and applied by Bhat (2003) and Train (2003). By re-ordering coefficients in the number generation rule, scrambling actually exchanges numbers' positions in the original standard Halton sequence (see, e.g., Bhat, 2003). As can be derived from the rule, the first several numbers in the sequence only exchange positions in their near neighborhood. As the sequence length increases, the exchanged positions are further apart, but then the values of those numbers will not differ by much. In other words, though scrambling can disrupt the correlation across dimensions (i.e., sequences generated with different prime numbers), within each scrambled Halton sequence, an unbroken correlation pattern remains. Compared to the original Halton sequence, a scrambled Halton sequence simply cycles in a different way.

Train's (2003) illustration helps the drawing rule of a scrambled Halton sequence. Still using Figure 1, a scrambled sequence is obtained by reversing the order of B and C. That is, the listing becomes: C, B, CA, CC, CB, BA, BC and BB. The positions of the first eight numbers in standard and scrambled Halton sequences can be shown as follows:

Standard Halton	3	6	1	4	7	2	5	8
Scrambled Halton	6	3	2	8	5	2	7	4

As can be inferred from this analysis, scrambled Halton sequences change number positions only slightly. When considering correlation across observations' simulated probabilities, where the ordering of numbers for a single observation does not have any influence, standard and scrambled Halton results will not differ much. In fact, with base 3, the correlation pattern generated by these two methods almost overlaps. Only with higher prime numbers does scrambling change sufficiently such that the difference between standard and scrambled sequences can be observed. Figure 4 shows the correlation across observations induced by a scrambled Halton sequence, with base 7. As shown, the sequence still generates a periodic correlation pattern, very similar to that emerging from a standard Halton sequence. In general, the peaks are slightly lower, but the maximum correlation (e.g., at $k = 7$ and its multiples) remains quite high.

Therefore, while scrambling will remove a standard Halton's correlation across high dimensions, it still has the same problem as a standard Halton sequence when it comes to inter-observation correlations.

The main features of a scrambled Halton sequence are representative of several other QMCs, including Sobol and Faure sequences. These two sequences remove correlation across high dimensions in a slightly different way (though within each dimension they work essentially the same as a Halton sequence). In other words, from the perspective of correlation, Sobol and Faure sequences function similar to scrambled Halton sequences. Thus, the scrambled Halton sequence is used here, as an example of this type of QMC simulation.

Shuffled Halton:

Hess and Polak (2003) first proposed to use the shuffled Halton sequence[2] in a mixed logit model. Compared to a scrambled sequence, the generation of a shuffled sequence is more straightforward: the same Halton sequence is randomly shuffled for different observations and alternatives. That is, if there are Q alternatives, and each has just one random coefficient and N observations, a total of $Q \times N$ short sequences need to be generated. If each sequence contains R draws, with a standard or scrambled Halton sequence, Q long sequences, each having $N \times R$ numbers, need to be generated. This means that with large N and R, these two approaches will require considerably large memory and computational time. With a shuffled Halton sequence, only one standard Halton sequence of length R needs to be generated. This short sequence is shuffled $Q \times N$ times.

One easy way to shuffle a sequence of numbers is to generate a vector of uniformly distributed random values and sort the sequence according to their order. There are $R!$ ways to permute a sequence containing R different numbers, and in most applications R is at least 50. Therefore, normally $Q \times N$ is much less than $R!$, making these $Q \times N$ sequences uncorrelated (see, e.g., Wang and Kockelman [2006]). This shuffling process can effectively disrupt correlations between alternatives. However, since this method uses the same sequence

[2] While Morokoff and Caflisch (1994) were the first to suggest a mathematically equivalent version of the shuffled Halton sequence, Hess and Polak (2003) developed their work independently and are the first one to use the term "shuffled Halton". They also were the first to apply this technique with a mixed logit model.

of numbers, the asymptotic properties may be problematic. In the perspective of correlation across simulated probabilities, a shuffled Halton sequence imposes *perfect* positive correlation across observations, as shown in Figure 4. This implies that shuffling a single Halton sequence for all observations is likely to cause a higher estimator variance through simulation noise. It should be less efficient than standard and scrambled Halton sequences (but not necessarily less efficient than PMC, thanks to its better coverage). When shuffling is used for correlated observations, this strong correlation pattern may obscure the true correlation patterns and harm prediction.

Therefore, shuffled Halton sequences should be used with some caution. The initial position and sequence length must be chosen carefully, in order to ensure that the sequence has optimal coverage and does not obscure any behavioral relationships of interest.

Long Shuffled Halton Sequence

In order to provide a more comprehensive comparison, two new methods are proposed and used here. The first one is referred to as "long shuffled Halton" sequence, because, when compared to Hess and Polak's shuffled sequence, the original Halton sequence used here has a length $N \times R$. The whole long sequence is first shuffled, then divided into N segments for those N observations(instead of first dividing and then shuffling, because as discussed previously, reordering a sequence within an observation does not influence the simulated probabilities. This means if a sequence is first cut and then shuffled, in terms of observation correlation, it is equal to using the original sequence and the correlation between simulated probabilities is still not broken.) Therefore, its coverage of each observation's distribution is not as uniform as that of a standard or scrambled Halton sequence. However, correlation across observations is low, as shown in Figure 4. By construction, this long shuffled Halton sequence should perform like PMC. The second method is a randomized shuffled sequence, as described below.

Randomized (Shifted) Shuffled Halton Sequence

In the second method, a shuffled Halton sequence is randomized, through shifting. A random number (drawn from a standard uniform distribution) is added to each shuffled Halton sequence. (If the resulting value exceeds 1.0, a value of 1 is subtracted.) As Bhat (2003) and Train (2003) describe, this operation preserves the sequences uniform distribution in the number space.

As noted earlier, shuffling offers good coverage within each observation. Its problem is the repeated usage of the same sequence of numbers. By adding different random numbers to different sequences (though the same random number is used within one sequence), the problems of shuffling are avoided. As shown in Figure 4, the observation correlation generated by this method is close to zero. Meanwhile, this method maintains good coverage. With all these characteristics, this method is expected to outperform PMC and all QMC methods discussed above, especially when observations are correlated. And this is indeed the case shown here. In the following sections, all five methods of number generation, together with PMC, are applied to a synthetic dataset for MSLE.

CORRELATED SYNTHETIC DATA

A synthetic dataset can be developed for any context, as long as the data-generating process is not internally inconsistent. Here, in order to make the correlation pattern more meaningful, the synthetic data may be interpreted as a case of land development. In this example, each observation stands for a grid cell of land at a specific location. There are totally 1500 observations, composing a rectangular area (shown in Figure 5). The value of the dependent variable y may be interpreted as the land use type. There are three alternatives: 1 indicates residential use, 2 indicates commercial use, and 3 indicates industrial use.

Observation i 's random utility for alternative q is specified as follows:

$$U_{iq} = \beta_{COST} \cdot COST_{iq} + \beta_{REV} \cdot REVENUE_{iq} + \left(\alpha_q + \sigma_q \eta_{iq}\right) + \varepsilon_{iq} \tag{6}$$

$COST_{iq}$ is generated from a uniform distribution with mean 1 and standard deviation 0.577. In a land development example, $COST_{iq}$ can be interpreted as the general cost of developing a grid cell, and is affected by zoning regulations and construction costs. β_{COST} is a fixed coefficient set to -3. $REVENUE_{iq}$ is drawn from a uniform distribution (mean 1.5 and standard deviation 0.866). It may be interpreted as a property's revenues, as determined by floorspace prices and/or rents. Its coefficient β_{REV} is fixed to equal 2.

There also is a variable ASC-component to the random utility function, that is normally distributed with mean α_q and standard deviation σ_q. This component can be interpreted as a constant term plus one normally distributed unobserved error term. Alternative one's constant term α_1 is -1, and its standard deviation σ_1 is 10. Alternative 2 has α_2 =1 and σ_2 =5. Alternative 3 is the base alternative and thus has corresponding parameters of 0 and 0. This example only considers correlation across observations: η_{iq} is uncorrelated across alternatives, but within one alternative, it follows a multivariate normal distribution with a mean vector of zeros and unit variance terms. In the first of the two test data sets, there is no correlation across these terms. In the second data set, the covariance terms, $\mathrm{cov}\left(\eta_{iq}, \eta_{jq}\right)$, are specified to depend on Euclidean distances between observations i and j, as described below. This is the genesis of dataset 2's spatial nature.

As is common in kernel logit models, ε_{iq} follows a standard Gumbel distribution with location parameter equal to 0 and scale parameter equal to 1. It is uncorrelated with the two explanatory variables and η. The total random term per observation and alternative is therefore $\sigma_q \eta_{iq} + \varepsilon_{iq}$. The covariance structure of η_{iq} produces spatial correlation in the random utilities for land use development. Thus, all told, there are six parameters requiring estimation: β_{COST}, β_{REV}, α_1, σ_1, α_2 and σ_2.

For each observation, the alternative with maximum utility becomes the outcome. Figure 5's first graph shows the land use outcomes when the η_{iq} terms are independent across observations (dataset 1). For dataset 2 (Figure 5's second illustration), a predetermined correlation pattern is assumed to exist across $\eta_{\bullet p}$ values : The correlation between two observations is calculated as $\exp\left(-dis_{ij}/10\right)$, where dis_{ij} is the Euclidean distance between the centroids of observations i and j, standardized by the length of one unit's edge. Thus, the correlation matrix is of size 1500×1500, and the correlation of η values between immediate neighbors is around +0.9. The correlation between observations i and j is

$$corr\left(U_{iq}, U_{jq}\right) = \frac{\sigma_q^{\,2}\,\mathrm{cov}\left(\eta_{iq}, \eta_{jq}\right)}{\sigma_q^{\,2} + \pi^2/6} \tag{7}$$

This means that for alternative 1, the correlation between adjacent observations (grid cells) is around +0.89; for alternative 2, it is around +0.85. (These differ slightly due to differing variance assumptions on their error components.) The correlation here seeks to capture correlations in unobserved attitudes and attributes related to developing parcels within neighborhoods.

RESULTS ANALYSIS

The five simulation methods discussed above, along with PMC, were applied to the synthetic data. Each technique is used seven times, with the number of draws (*R*) varying from 5 to 50, 100, 200, 500, 1000 and 2000. In order to avoid potential identification problem (Ben-Akiva et al., 2001), the fixed coefficient β_{COST} was constrained to equal its true value while all other parameter estimates were allowed to change during the MSLE process. All starting values were set to their true values, so that problems of local extreme are avoided. For QMCs and their hybrid methods, the bases used to generate $\eta_{i,1}$ and $\eta_{i,2}$ were 3 and 5, respectively.

1,500 observations is not a very large sample size (As can be inferred above, *N* systematically correlated observations lead to a correlation matrix of size $N \times N$. Thus a larger sample size will make the above synthetic data generation practically infeasible with its excessive computational burden.) Hence, data sampling variance and bias may contribute to differences observed here between estimates and the datasets' true values. To account for this, the study uses 10 samples (or 20 data sets total), each drawn from the same distribution described above (with 10 exhibiting spatial independence and 10 exhibiting spatial correlation in random error components). Average values of estimates from these samples were compared to their true values. Since each model requires 5 parameter estimates, to illustrate the overall accuracy, the roots of the sum of squared errors (RSE) of these five parameters were calculated. This measure can be interpreted as the Euclidean distance between the vector of estimated values and the vector of true values: $\sqrt{\left\|\hat{\beta}-\beta\right\|^2}$. From this perspective, this measure serves as a generalized bias.

Table 1 shows this generalized estimate of bias with independent observations (i.e., dataset 1). As expected, standard and scrambled Halton sequences outperform PMC. With *R*=50, the simulation bias is already quite low using either of these two methods, while PMC does not achieve equivalently low bias until *R* exceeds 200. Also as expected, the shuffled Halton sequence performs better than PMC but not as well as standard and scrambled Halton sequences. The long shuffled Halton sequence performs much like PMC, in terms of trend, maximum and minimum biases. The randomized shuffled Halton does not produce good estimates with low R values (e.g., *R*=50); however, once R exceeds 100, this method yields a smaller generalized bias than PMC, shuffled Halton and long shuffled Halton. The overall findings here confirm conclusions from previous studies, in that scrambled Halton sequences perform better than shuffled and other Halton sequences when nothing is amiss in the data series. In addition, it supports the notion that randomized shuffled Halton sequences yield fairly satisfactory results.

Table 2 shows the generalized bias values for dataset 2's correlated observations. With this example, the 6 methods of MSLE sequence generation yield almost the same results. This result suggests that concerns relating to observation correlation generated by QMCs may be unnecessary. However, it should be noted that the spatial correlation pattern inherent in dataset 2 does not differ too much from the patterns generated by standard, scrambled and shuffled Halton sequences: The correlation between error terms of nearby observational units in space is always positive, paralleling the shuffled Halton's perfect positive correlation pattern. The rule for generating this synthetic data also suggests that the correlation peaks somewhere between every 30th and 50th observation pairing. These are multiples of 3 and 5, which were used as the bases for all five types of Halton sequence used here. If the synthetic data set has a different pattern and it is completely counter to the sequences, the pattern falsely imposed by sequences may obscure the real data correlation pattern and affect estimators, especially those associated with the variance covariance matrix. Therefore, if Halton sequences must be used, a safe approach may involve shuffling the data prior to analyzing it, so that the potential correlation can be removed (unless, of course, the order of observations is important to the model specifications, as in serial or spatial correlation analysis, where estimation methods are much facilitated by proper ordering of observations).

As to the performance of the randomized shuffled halton sequence, thanks to the benefits of Halton sequence coverages, only when R=1000 does PMC perform better. With all other draw numbers, the randomized shuffled Halton sequence appears to yield smaller errors. The results from this experiment suggest that, even when observations are correlated, QMCs and hybrid methods may be preferred to PMC, thanks to their better coverage.

Table 3 shows estimates from just one of the 10 samples. For this sample, the estimates that result from assuming fixed coefficients on the *COST* and *REVENUE* terms are more accurate than those for the true random coefficients. These results also suggest that estimates from standard and scrambled Halton sequences are the most stable as R increases. With empirical datasets, where true parameter values are not known, one signal for convergence, as Hensher and Greene (2003), Train (2003) and Walker (2001) have all suggested, is the stabilization of estimate values. Therefore, this feature of standard and scrambled Halton sequences also makes them preferable for practice.

CONCLUSIONS

With the growth of computational capability in recent years, simulation has been broadly adopted for complex model estimation. Significant effort has been devoted to developing more efficient simulation techniques. This paper discusses the advantages and potential problems of several simulation techniques, including standard, scrambled, shuffled and long shuffled Halton and randomized shuffled Halton sequences. Standard sequences are found to generate periodic correlations across observations, with cycle lengths equal to their bases (or multiples of their bases, for higher dimensions). Though a scrambled Halton sequence can effectively disrupt a standard Halton's high correlation across alternatives, it cannot break the correlation across observations. A shuffled Halton sequence uses the same sequence of numbers for each observation, thus imposing perfect correlation across observations.

This issue of observation correlation deserves some attention, because in some spatially, temporally, or otherwise correlated data sets and models, there may be no alternatives for multi-dimensional integration. In these cases, there will be no need to consider shuffling or scrambling, in order to break correlations across alternatives (in a model of discrete choices across alternatives). However, the inter-observation correlation caused by a standard Halton sequence may be just as much of an issue: The correlation patterns across observations, when significantly different from the data set's true correlation patterns, may increase the simulation variance of the likelihood function and resulting estimates. For finite samples, this also may contribute to a situation of empirical unidentifiability, where the true log likelihood function is already flat, due to large sampling variance.

In this work, a standard Halton, scrambled Halton, shuffled Halton, long shuffled Halton, and randomized shuffled Halton, together with PMC sequences were applied to a synthetic data set where random coefficients correlated across observations, thanks to spatial processes. Fortunately, the results suggest that when the generic and imposed correlation patterns are not at odds, these QMC and hybrid methods are at least as good as PMC, thanks to their better coverage.

Furthermore, the randomized shuffled Halton sequence was found to perform fairly well, with both independent and correlated observations. Considering that this method requires less computational time and memory than the scrambled sequence, this hybrid method may serve as a good alternative to standard and scrambled Halton sequences. Finally, it should be noted that, as realized in many prior studies, properties of a QMC sequence depend on its base, number of draws, and number of observations, as well as its initial position. Our knowledge of QMC is still quite limited, so these newly developed techniques should be used with caution. When the model is complicated and there is reason to believe potential observation correlation exists, PMC may be the safest choice.

Many opportunities for further study exist in this domain, including variations on inter-observational dependence and comparisons of Bayesian and classical (MSLE) approaches for estimation of these complex models. As Bolduc et al. (1997) suggests, a Bayesian procedure requires about half as much computer time as MSLE with PMC. It would be interesting to see how Bayesian simulation techniques perform with correlated observations, and how results compare to those based on QMC sequences for MSLE.

ACKNOWLEDGMENTS

This work is financially supported by the Benjamin H. Stevens Graduate Fellowship in Regional Science, provided by the North American Regional Science Council. The authors are grateful to those who provided help for this study. These include Kenneth Train for his inspiring emails and Gauss code for estimating mixed logit models, Olivier Teytaud for his Matlab code generating scrambled Halton sequences, Joan Walker for her email replies, and Annette Perrone for her administrative and editing assistance.

REFERENCES

Ben-Akiva, M., D. Bolduc, and J. Walker (2001) Specification, estimation and identification of the logit kernel (or continuous mixed logit) model. (Working Paper). Department of Civil Engineering, MIT. [Accessed June 10, 2006: http://emlab.berkeley.edu/reprints/misc/multinomial2.pdf]

Bhat, C. (2003) Simulation estimation of mixed discrete choice models using randomized and scrambled Halton sequences. *Transportation Research 37B* (9): 837-855.

Braaten, E. and G. Weller (1979) An improved low-discrepancy sequence for multidimensional quasi-Monte Carlo integration. *Journal of Computational Physics* 33 (2): 249-258.

Bolduc, D., B. Fortin, and S. Gordon (1997) Multinomial probit estimation of spatially interdependent choices: An empirical comparison of two new techniques. *International Regional Science Review* 20 (1-2): 77-101.

Faure, H. (1992) Good permutations for extreme discrepancy. *Journal of Number Theory* 42 (1): 47-56.

Greene, W. (2002). *Econometric Analysis, 5th Edition*. Upper Saddle River: Prentice-Hall.

Halton, J. (1960) On the efficiency of evaluating certain quasi-random sequences of points in evaluating multi-dimensional Integrals. *Numerische Mathematik* 2: 84-90.

Hellekalek, P. (1984) Regularities in the distribution of special sequences. *Journal of Number Theory* 18 (1): 41-55.

Hensher, D. and W. Greene (2003) The mixed logit model: the state of practice. *Transportation* 30 (2): 133-176.

Hess, S. and J.W. Polak (2003) The Shuffled Halton sequence. (Working Paper). Centre for Transport Studies, Imperial College London. [Accessed June 10, 2006: http://www.cts.cv.imperial.ac.uk/StaffPages/StephaneHess/papers/Hess_and_Polak_Shuffled_Halton_Sequences.pdf]

Morokoff, W.J. and R.E. Caflisch (1994). Quasi-random sequences and their discrepancies. *SIAM Journal of Scientific Computing* 15 (6): 1251-1279.McFadden, D. (1989). A method of simulated moments for estimation of the multinomial probit without numerical integration. *Econometrica* 57(5): 995-1026.

McFadden, D. and K. Train (2000). Mixed MNL models of discrete response. *Journal of Applied Econometrics* 15(5): 447-470.

Munizaga, M. and R. Alvarez-Daziano (2001) Mixed logit versus nested logit and probit. (Working Paper). Department of Civil Engineering, The University of Chile. [Accessed June 10, 2006: http://tamarugo.cec.uchile.cl/~dicidet/mmunizaga/mixed_logit.pdf]

Niederreiter, H. (1992) Random number generation and quasi-Monte Carlo methods. Presented at CBMS-NSF Regional Conference Series in Applied Math, SIAM, Philadelphia, Pennsylvania.

Rouwendal, J. and E. Meijer (2001) Preferences for housing, jobs, and commuting: A mixed logit analysis. *Journal of Regional Science* 41(3), 475-505.

Sobol, I. M. (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Comput. Maths. Math. Phys.*, 7, 86-112.

Spanier, J. and E.H. Maize (1994) Quasi-random methods for estimating integrals using relatively small samples. *SIAM Review* 36(1), 18-44.

Srinivasan, K. and H. Mahmassani (2005). Dynamic kernel logit model for the analysis of longtitude discrete choice data: Properties and computational assessment. *Transportation Science* 39 (2): 160-181.

Train, K. (1999) Halton sequences for mixed logit. (Technical Paper). Department of Economics, University of California, Berkeley. [Accessed June 10, 2006: http://129.3.20.41/eps/em/papers/0012/0012002.pdf]

Train, K. (2003) *Discrete Choice Methods with Simulation*, Cambridge University Press, New York.

Walker, J. L. (2001) *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures, and Latent Variables*. Ph.D. Dissertation, Department of Civil and Environmental Engineering, MIT.

Wang, X. and K. M. Kockelman (2006) Tracking land cover change in a mixed logit model: recognizing temporal and spatial effects. *Transportation Research Record* No. 1977: 112-120.

LIST OF TABLES

LIST OF FIGURES

FIGURE 4 Observation correlation generated by different simulation techniques (base=7, R=100).

FIGURE 5 Outcomes of synthetic data.

Table 1. Generalized Bias (for All Parameters) with Independent Observations (Dataset 1)

Number of Draws (R)	PMC	Standard Halton	Scrambled Halton	Shuffled Halton	Long Shuffled Halton	Randomized Shuffled Halton
5	6.333	9.803	1.900	4.742	6.686	4.380
50	0.451	0.190	0.156	0.310	0.472	0.524
100	0.286	0.223	0.132	0.389	0.275	0.208
200	0.224	0.153	0.167	0.167	0.158	0.154
500	0.174	0.151	0.144	0.162	0.267	0.161
1000	0.161	0.159	0.160	0.184	0.163	0.175
2000	0.196	0.163	0.161	0.180	0.158	0.139

Table 2. Generalized Bias (for All Parameters) with Correlated Observations (Dataset 2)

Number of Draws (R)	PMC	Standard Halton	Scrambled Halton	Shuffled Halton	Long Shuffled Halton	Randomized Shuffled Halton
5	6.998	2.776	2.550	8.443	7.145	2.786
50	1.563	1.310	1.294	1.417	1.681	1.404
100	1.430	1.452	1.425	1.143	1.555	1.368
200	1.555	1.380	1.341	1.398	1.366	1.436
500	1.521	1.416	1.427	1.344	1.420	1.361
1000	1.396	1.384	1.400	1.446	1.327	1.468
2000	1.429	1.407	1.409	1.395	1.399	1.385

Table 3. Parameter Estimates with Different Methods and Datasets

Method	No. of Draws (R)	Dataset 1 (Independent Observations)					Dataset 2 (Correlated Observations)				
	True Value	β_{REV}	α_1	σ_1	α_2	σ_2	β_{REV}	α_1	σ_1	α_2	σ_2
		2	-1	10	1	5	2	-1	10	1	5
PMC	5	0.988	0.223	5.166	0.710	1.629	1.360	-0.390	3.283	0.848	1.603
	50	1.722	-0.975	9.408	0.888	4.986	1.990	-2.127	6.924	1.043	3.408
	100	1.737	-1.045	9.427	0.837	4.962	2.115	-2.253	7.218	1.096	3.606
	200	1.847	-1.136	9.917	0.781	5.074	2.108	-2.289	7.231	1.079	3.597
	500	1.828	-1.064	9.863	0.816	4.962	2.156	-2.177	7.355	1.092	3.643
	1000	1.818	-1.059	9.693	0.796	5.003	2.141	-2.152	7.120	1.077	3.624
	2000	1.824	-1.107	9.810	0.796	4.940	2.128	-2.145	7.123	1.090	3.605
Standard Halton	5	1.260	-1.475	9.350	0.536	3.406	1.728	-1.047	5.205	0.847	3.056
	50	1.815	-0.864	9.115	0.729	5.076	2.146	-2.096	7.070	1.066	3.684
	100	1.851	-1.120	9.931	0.787	5.013	2.112	-2.176	7.102	1.084	3.585
	200	1.835	-1.122	9.882	0.804	5.007	2.137	-2.102	7.023	1.085	3.665
	500	1.828	-1.099	9.813	0.804	4.964	2.135	-2.149	7.110	1.086	3.633
	1000	1.836	-1.123	9.903	0.811	4.985	2.133	-2.155	7.113	1.087	3.624
	2000	1.832	-1.114	9.853	0.811	4.980	2.133	-2.155	7.113	1.087	3.622
Scrambled Halton	5	1.392	-0.587	7.871	0.482	5.029	1.716	-1.202	5.275	0.846	3.421
	50	1.871	-1.048	9.689	0.770	5.223	2.108	-2.117	7.037	1.048	3.677
	100	1.820	-1.083	9.783	0.813	4.954	2.130	-2.141	7.115	1.085	3.657
	200	1.847	-1.115	9.945	0.808	5.043	2.126	-2.187	7.196	1.086	3.600
	500	1.838	-1.124	9.935	0.812	4.976	2.136	-2.138	7.071	1.086	3.642
	1000	1.830	-1.093	9.796	0.807	4.991	2.136	-2.164	7.139	1.089	3.621
	2000	1.833	-1.119	9.876	0.812	4.979	2.132	-2.156	7.116	1.087	3.622
Shuffled Halton	5	1.344	-3.118	12.545	0.804	3.505	1.863	-9.976	20.681	1.109	3.117
	50	1.813	-1.119	9.997	0.953	5.084	2.075	-2.364	7.639	1.236	3.493
	100	1.785	-0.965	9.550	0.820	4.996	2.104	-2.235	7.353	1.120	3.621
	200	1.820	-1.058	10.055	0.911	4.908	2.113	-2.059	7.034	1.141	3.636
	500	1.825	-1.110	9.972	0.843	4.939	2.129	-2.220	7.332	1.108	3.605
	1000	1.826	-1.104	9.932	0.847	4.959	2.128	-2.166	7.175	1.107	3.630
	2000	1.837	-1.219	10.198	0.847	4.953	2.137	-2.164	7.165	1.103	3.617
Long Shuffled Halton	5	0.946	0.012	4.101	0.690	2.240	1.333	-0.654	3.473	0.904	1.421
	50	1.841	-0.797	9.197	0.614	5.447	2.087	-2.260	7.441	1.055	3.587
	100	1.834	-1.093	10.026	0.793	5.354	2.053	-2.104	7.028	1.076	3.515
	200	1.822	-1.044	9.476	0.770	4.997	2.126	-1.960	6.751	1.054	3.650
	500	1.805	-1.236	10.070	0.821	4.775	2.108	-2.249	7.290	1.098	3.535
	1000	1.823	-1.112	9.871	0.819	4.950	2.152	-2.171	7.127	1.096	3.653
	2000	1.837	-1.019	9.592	0.790	5.031	2.144	-2.163	7.131	1.094	3.634
Randomized Shuffled Halton	5	1.152	-2.470	14.995	0.747	2.742	1.597	-1.786	6.586	0.891	2.183
	50	1.732	-1.788	11.769	0.884	4.764	2.078	-2.124	7.190	1.080	3.545
	100	1.793	-1.182	9.957	0.827	4.929	2.093	-2.108	7.081	1.070	3.556
	200	1.799	-1.101	9.687	0.781	5.026	2.114	-2.258	7.285	1.110	3.529
	500	1.823	-1.115	9.934	0.821	4.893	2.138	-2.180	7.114	1.094	3.612
	1000	1.828	-1.098	9.815	0.804	4.975	2.123	-2.165	7.103	1.082	3.605
	2000	1.834	-1.157	9.941	0.822	4.965	2.135	-2.140	7.113	1.090	3.615

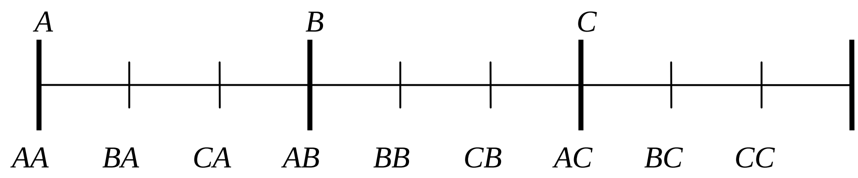


Figure 1. Segments for standard and scrambled Halton sequences (from Train (2002)).

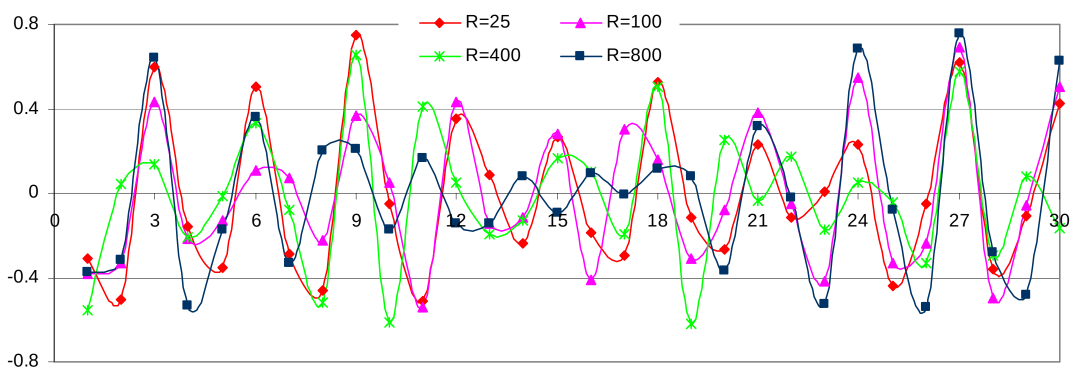


Figure 2. Observation correlation generated by Halton sequence with different numbers of draws (base=3).

Note: Y = $corr\left(SP_{\bullet}, SP_{\bullet+k}\right)$, X = k

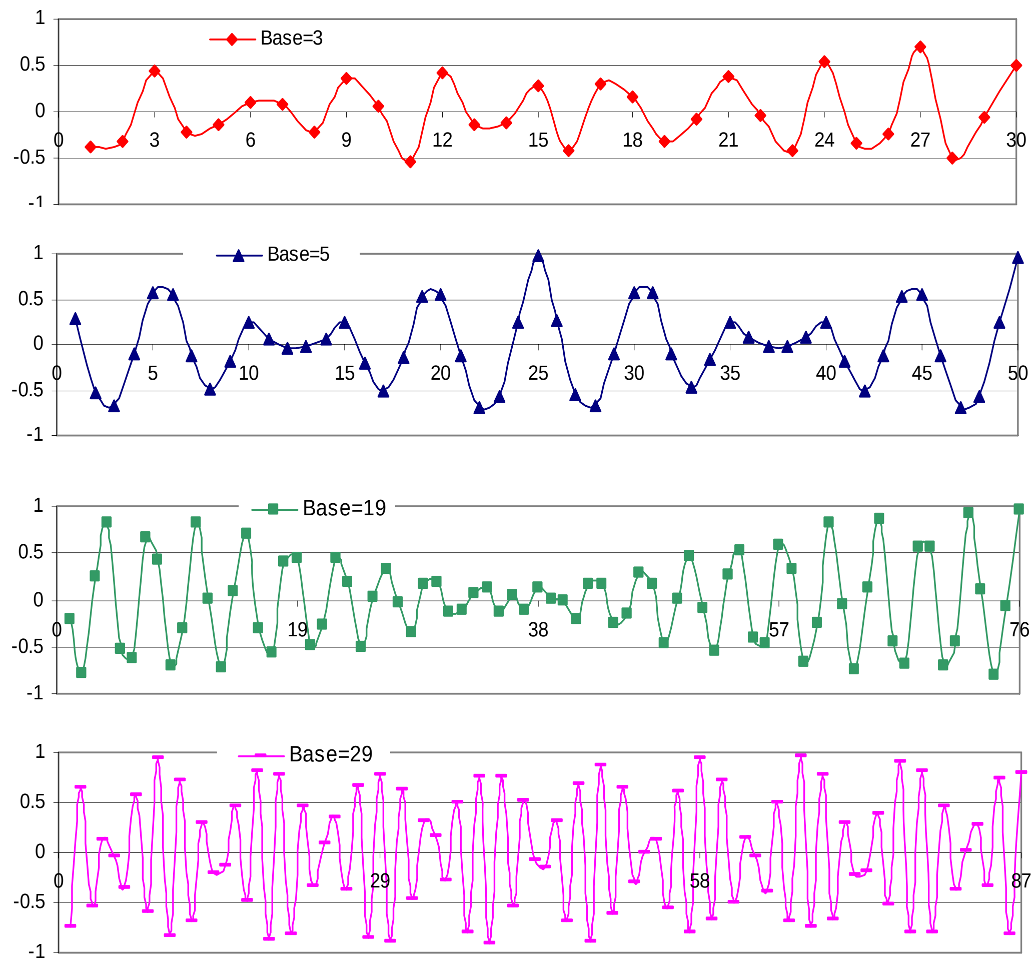


Figure 3. Observation correlation generated by Halton sequences with different bases (*R*=100).

Note: Y = $corr\left(SP_{\bullet}, SP_{\bullet+k}\right)$, X = k

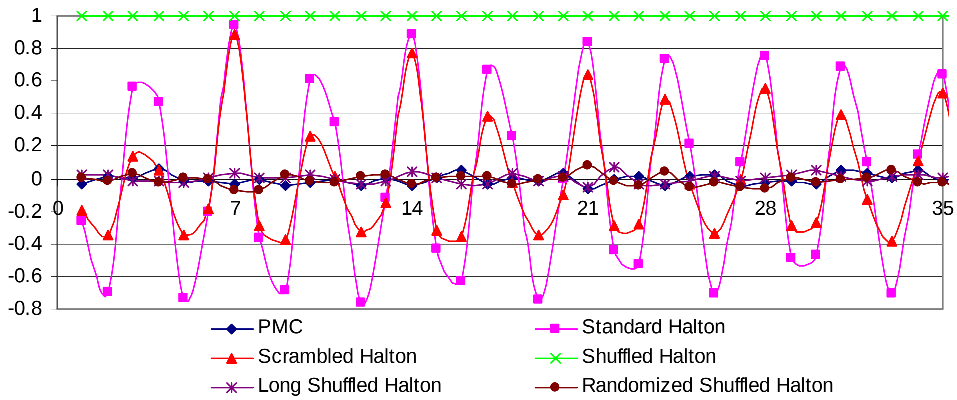
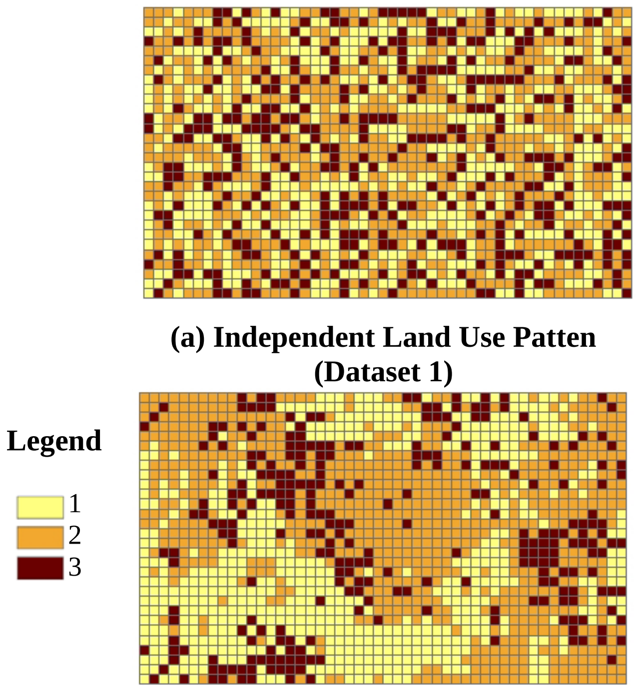


Figure 4. Observation correlation generated using different simulation techniques (base=7, *R*=100).

Note: Y = $corr\left(SP_{\bullet}, SP_{\bullet+k}\right)$, X = k



(a) Independent Land Use Patten (Dataset 1)

(b) Correlated Land Use Pattern (Dataset 2)

Figure 5. Outcomes of the synthetic data.