Multi-Dimension and Real-Time Interpolation
(Dirac-Monte Carlo Method)

K.K. (Benjamin) Fang

[*] *FANG, INC., Stanton, CA 90680 U.S.A.*

E-mail: fanginc@gte.net

Abstract: A new interpolation method is used for Spatial Interpolation Comparison (SIC) 2004 exercise. The new interpolant is constructed by use of Dirac delta function and Monte-Carlo (integration) method. The interpolant is dependent upon "coordinate separation" rather than "distance". Mathematical treatment of the Dirac-Monte Carlo (DMC) method has been given in detail in this paper. Interpolation calculations were done with two input files provided for the exercise. Comparison between the calculated and the true radioactivity measurement was made. Error analysis and uncertainty, together with measure-of-merit had been given. Contour maps for the estimations are presented for both input files. New and future development of DMC method were discussed. [Description of SIC 2004 exercise and data set used can be found at: http://www.ai-geostats.org/events/sic2004/]

Keywords: Multi-dimension interpolation, real-time computation, Dirac delta function, Monte-Carlo multiple integration, estimation errors.

1. INTRODUCTION

A formulation of new interpolation method, Dirac-Monte Carlo method (DMC), is presented in this paper. Though DMC method produces interpolants in arbitrary dimension (1, 2, 3,..10,...100,....) in a straightforward manner, interpolant in terms of Cartesian coordinates for two-dimension space is described in detail in this presentation for SIC2004 exercise.

DMC method was published at *Random Data Interpolation Center* (RDIC) in 2002 by our organization (FANG, INC.) on the web at the following address: http://www.fanginc.com/main.htm and it has been continually updated ever since. RDIC is dedicated to solve interpolation problems with given input data at random locations. As the title of this paper indicates, DMC method is aimed at providing solutions for multi-dimension and real-time applications. In this context, we are delighted to participate in SIC2004 exercise to demonstrate the merits of DMC method. There are a few aspects of DMC method that need to be pointed out up front. The method does not perform any "data dependent" analysis (no variogram study) prior to the actual interpolation in contrast to Kriging and it does not give emphasis to the process of "detection of outliers and anomalies" either. The method honors every input function value and is strictly based on mathematics and statistics. DMC interpolant is application independent, and the interpolation results produced are numerically stable and statistically assured. Furthermore, the method does not employ transcendental functions nor any computing intensive procedures (such as root finding, curve fitting, optimization, etc.). Hence, the method is particularly helpful in cutting down the computing time and generating answers in real-time which is of course vital to deal with emergency situation in the environment, for example, surge of radioactivity (same like SIC 2004 joker data set), biological agent or chemical substance. Finally, on the operation side of the method, it needs input value for “delta width” or “kernel bandwidth” (to be explained in the following section) for each dimension. For SIC 2004 exercise, it needs two delta width values, one for x-axis and one for y-axis. DMC method provides first-cut value for delta width but the value can be and should be fine tuned based upon the practitioner’s expertise on the problem at hand.

2. METHODOLOGY

It is challenging to construct “interpolant” which can be deduced through mathematical analysis by use of given input function values at random locations. The well-known Lagrangian interpolation formula (ref. 1) in one-dimension is still in use today. However, the Lagrangian formula can not be shown through mathematical derivation. Therefore, it is not straightforward to extend the formula to 2-dimension and above. Popular interpolation methods (Shepard’s distance-weighted, Hardy’s multiquadrics, Kriging, etc.) have been developed and practiced successfully in the past few decades in different industries, as well as in scientific and engineering research (refs 2, 3). On the other hand, in statistics community the so-called nonparametric kernel regression (ref. 4) has been studied in the past 50 years, and many analytical results have been discovered, including interpolants (called estimators) and their associated errors and convergence rates. In fact, the original Shepard’s interpolant looks somewhat similar to the well-known “Nadaraya-Watson estimator” practiced in kernel regression analysis. Starting from a different approach through our observation, we are able to derive analytically, and establish quickly the interpolant formula for 1-dimension, 2-dimension and any higher dimension. The interpolant found in Dirac-Monte Carlo method has been identified and it is closely related to Nadaraya-Watson estimator. However, one distinct feature of DMC interpolant, different from other interpolants/estimators, is that DMC interpolant is dependent upon individual “coordinate separation”, not on the “distance”. This difference makes DMC interpolant capable of handling non-convex domain (For example, in between two concentric spherical shells in 3-D or two concentric circles in 2-D, or L-shape corridor.). With the help of Dirac delta function, it is straightforward to generalize DMC interpolants in terms of non-Cartesian coordinates, such as polar coordinates, spherical coordinates, cylindrical coordinates, etc. (ref. 5). Furthermore, the uncertainty analysis of DMC interpolant is derived directly through the use of Central Limit Theorem and radically different from the findings of kernel regression method. Due to the fact that DMC is a new interpolation method, we present the mathematical analysis below to describe DMC method. (Please also view web pages, including references, FAQs, and comparison with other intepolants provided at RDIC)

First, the two ingredients, *Dirac delta function* and *Monte-Carlo method*, used in the formulation are presented:

(1) Dirac delta function (Refs. 6, 7)

Dirac delta function is a special impulse, weighting function which has the following properties:

$$f(x) = \int_a^b f(x')\,\delta(x'-x)\,dx' \text{ ; where a<x<b} \qquad \text{(Eq. 1)}$$

$f(x)$ is continuous and bounded;

δ(x'-x) = 0 when x' not equal to x

δ(x'-x) = ∞ (infinity) when x' equal to x

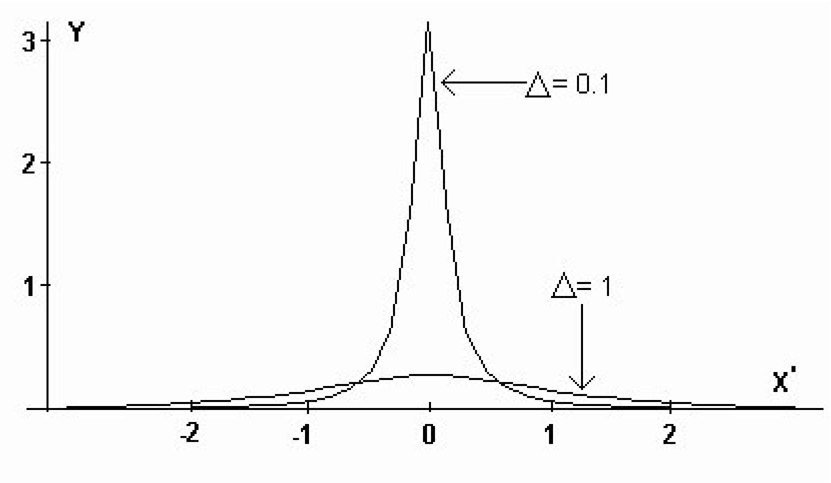
$$\int_{-\infty}^{+\infty} \delta(x'-x)dx' = 1 \text{ ;} \qquad \text{(Eq. 2)}$$

Note that the delta function was introduced by Paul Dirac in theoretical physics in the early 20th century. The function has been given a rigorous, mathematical treatment by L. Schwartz with theory of Distribution. A number of analytical expressions of delta function are commonly used in the literature. They exist in the forms of rational function, transcendental function and infinite series expansion. The rational function below will be used to represent Dirac delta function in our work.

$$\delta(x'-x) = \frac{1}{\pi}\left[\frac{\Delta}{(x'-x)^2+\Delta^2}\right] \qquad \text{(Eq. 3)}$$

where the width Δ is a small quantity. It should be said that the rational function in (Eq. 3) is known as "Lorentzian" or "Breit-Wigner distribution" in physics and widely used for "resonance" phenomena in classical mechanics and quantum mechanics. It is also known as "Cauchy density" in statistics. A graphical display of delta function (Eq. 3) is shown in Figure A, where Δ =0.1 and Δ =1.0 with x=0. The abscissa is x' axis and the ordinate Y is Dirac delta function.

Figure A. Dirac Delta Function



Note that the smaller Δ is, the higher the peak will be and the more narrow the peak will become. Dirac delta function is peaked symmetrically and drops quickly towards zero value. The function has half peak value when x'-x = $\pm\Delta$. The area under the curve is always equal to 1 according to Eq. 2. In theory, it demands that the width approaches zero. However, in numerical computation, it can be set to a small (relative to the domain interval) and finite number. The delta width is an important parameter and will be discussed further later in the section.

(2) Monte-Carlo Integration (ref. 8 and ref. 9)

In the late 1940's, a novel technique was developed by E. Fermi, J. von Neumann, and S. Ulam in the area of evaluating integrals numerically. The method has been proven a powerful tool to handle computations of multi-variable problems in diverse subjects, physics, chemistry, biology, economics etc. In particular, Monte-Carlo method has been a great help to numerically evaluate multiple integrals in applications. There are two fundamental theorems behind Monte-Carlo method: (a) Strong law of large numbers; (b) Central limit theorem. They are briefly stated below to facilitate the presentation of Dirac-Monte Carlo method.

(a) Strong law of large numbers

If a sequence of N random variables x_1 to x_N are picked from a population with the probability density function g(x) and a new random variable A defined by the equation,

$$A=\frac{1}{N}\sum_{i=1}^{N} Z(x_i)\,, \qquad \text{(Eq. 4)}$$

where Z(x) is a given integrable function, and if the integral

$$\overline{Z}=\int_{-\infty}^{+\infty} Z(x)g(x)dx \qquad \text{(Eq. 5)}$$

exists, then A, with probability 1, approaches $\overline{Z}$ as a limit as N approaches infinity.

(b) Central limit theorem

For large N, the probability density distribution of A, G(A), is Gaussian, centered at $\overline{Z}$ with a standard deviation ($\frac{1}{\sqrt{N}}$) times that of the distribution of Z,

$$G(A)\xrightarrow[N\to\infty]{}\frac{1}{\sqrt{2\pi}\,(\frac{\sigma}{\sqrt{N}})}\exp[-\frac{(A-\overline{Z})^2}{2(\frac{\sigma}{\sqrt{N}})^2}] \qquad \text{(Eq. 6)}$$

where σ is the standard deviation of Z.(That is, $\sigma^2 = \overline{(Z-\overline{Z})^2}$). The above result is independent of the nature of Z(x) or g(x). In essence, the probability that the deviation of

A from $\overline{Z}$ will exceed $\pm\frac{\sigma}{\sqrt{N}}$ is 31%, $\pm\frac{2\sigma}{\sqrt{N}}$ 4.5%, $\pm\frac{3\sigma}{\sqrt{N}}$ 0.3%.

Now, the formulation of Dirac-Monte Carlo (DMC) interpolation is described below. We observe that the following equation exists,

$$\int_{a_1}^{b_1} [f(x') - f(x)]\ \delta(x' - x)dx'=0 \qquad \text{(Eq. 7)}$$

where x is the arbitrary value of x' variable and $a_1<x<b_1$, $f(x')$ is continuous and $\delta(x'-x)$ is the Dirac delta function. Next, using the density function defined as,

$g(x') = 1/(b_1-a_1)$, for $a_1 \le x' \le b_1$, and $g(x') = 0$, otherwise; (Eq. 8)

(Eq. 7) is recast by use of (Eq. 4) and $Z(x) = (b_1-a_1)\ [f(x') - f(x)]\delta(x'-x)$, $M'=N$, and it gives,

$$\frac{(b_1-a_1)}{M'}\sum_{i=1}^{M'}[f(x_i)-f(x)]\delta(x_i-x)\approx 0 \qquad \text{(Eq. 9)}$$

and

$$f(x)\sum_{i=1}^{M'}\delta(x_i-x)\approx\sum_{i=1}^{M'}[f(x_i)]\delta(x_i-x)$$

Then, f_A(x)is defined as,

$$f_A(x) = \frac{\sum_{i=1}^{M'}[f(x_i)]\ \delta(x_i-x)}{\sum_{i=1}^{M'}\delta(x_i-x)} \qquad \text{(Eq.10)}$$

where x_i are randomly chosen in the interval (a_1, b_1). It can be seen that by providing x_i, and $f(x_i)$, (Eq. 10) can be used to interpolate the function f(x) at location x where $a_1<x<b_1$. Note that f_A(x) is the searched interpolant. The accuracy and the convergence of f_A(x) are governed

by the central limit theorem which gives $\frac{1}{\sqrt{M'}}$ dependence (See ref. 6). With higher M' value, f_A (x) will approach closer to f (x). It should be said that (Eq. 10) has the same form as the famous nonparametric "Nadaraya-Watson" kernel regression estimator, f_{NW} (x) which is defined as, (ref.10 an ref. 11)

$$f_{NW}\,(\mathrm{x}) \quad = \quad \frac{\sum_{i=1}^{M'}[f(x_i)]K_H(x-x_i)}{\sum_{i=1}^{M'}K_H(x-x_i)} \qquad \texttt{where} \quad K_H(x-x_i)=(\frac{1}{H})K(\frac{x-x_i}{H})$$

Compare the above equation with Eqs. 10 and 3, one obtains that $\delta(x_i - x) = \delta(x - x_i) =$

$K_H(x - x_i)$ and Δ = H. Furthermore,

$$K(\frac{x-x_i}{H}) = \delta(\frac{x-x_i}{\Delta}) = (\frac{1}{\pi})[\frac{1}{(x-x_i)^2+1}]$$

H factor is also called "width" or "band width" in nonparametric kernel regression and controls the kernel smoothing property. The connection between Dirac-Monte Carlo method and nonparametric kernel regression can be understood because Dirac delta function is defined as a special "local, weighting function".

For 2-dimension Cartesian space, (Eq. 7) and (Eq. 9) are generalized respectively to,

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1}[f(x_1',x_2')-f(x_1,x_2)]\delta(x_1'-x_1)\,\delta(x_2'-x_2)\,dx_1'dx_2'=0 \qquad \texttt{(Eq. 11)}$$

$$f_A(\mathrm{x}_1,\mathrm{x}_2) = \frac{\sum_{i=1}^{M'} f(x_{1i},x_{2i})\,\delta(x_{1i}-x_1)\,\delta(x_{2i}-x_2)}{\sum_{i=1}^{M'}\delta(x_{1i}-x_1)\,\delta(x_{2i}-x_2)} \qquad \texttt{(Eq.12)}$$

With the use of (Eq. 3), we rewrite (Eq. 12) as,

$$f_A(x_1, x_2) = \frac{\sum_{i=1}^{M'}\frac{[f(x_{1i},x_{2i})]}{[(x_{1i}-x_1)^2+\Delta_1{}^2][(x_{2i}-x_2)^2+\Delta_2{}^2]}}{\sum_{i=1}^{M'}\frac{1}{[(x_{1i}-x)^2+\Delta_1{}^2][(x_{2i}-x_2)^2+\Delta_2{}^2]}} \qquad \text{(Eq. 13)}$$

The above equation (Eq. 13) is the 2-dimensional interpolant which is used to perform the calculation for SIC 2004 exercise. Note that x_1 is the x-coordinate and x_2 is the y-coordinate. Thus, (x_{1i}, x_{2i}) is the ith input random location and $f(x_{1i}, x_{2i})$ is the associated natural ambient radioactivity. (x_1, x_2) is the requested location where radioactivity is to be calculated. $M^{'}$ is the total number of input locations which is 200 according to SIC 2004 data sets. With Δ_1 and Δ_2 values given (preset), (Eq. 13) can be used to calculate the interpolated function value, $f_A(x_1, x_2)$. It can be seen that the interpolant, (Eq. 13), depends on the "product"of two coordinate-separation terms,

$$[(x_{1i}-x_1)^2+\Delta_1{}^2]\,[(x_{2i}-x_2)^2+\Delta_2{}^2]$$

and not on the distance which is defined as the square root of $(x_{1i}-x_1)^2 + (x_{2i}-x_2)^2$. Due to the intrinsic character of random nature of Monte-Carlo method, the interpolant is particularly suitable to solve SIC 2004 exercise. It should be said that Δ_1 and Δ_2 values can be estimated within the framework of DMC. The following formula can be found (See ref. 6),

$$\frac{(b_1-a_1)(b_2-a_2)}{M'\pi^2\Delta_1\Delta_2} \approx 1 \qquad \text{(Eq. 14)}$$

By assuming $\Delta_1 = \Delta_2$, (Eq. 14) becomes

$$\frac{(b_1-a_1)(b_2-a_2)}{M'\pi^2\Delta_1^2} \approx 1 \qquad \text{(Eq. 15)}$$

One can easily calculate the value of Δ_1 when the interval length (b_1-a_1) for x-axis and (b_2-a_2) for y-axis are provided .

We now begin the analysis of accuracy of (Eq. 13) and present the error analysis in terms of the famous Central Limit Theorem. Recalling (Eq. 9) and changing the "approximate" sign to "equal" sign", we obtain

$$\frac{(b_1-a_1)}{M'}\sum_{i=1}^{M'}[f(x_i)-f(x)]\delta(x_i-x)=E$$

$$f(x)\sum_{i=1}^{M'}\delta(x_i-x)=\sum_{i=1}^{M'}[f(x_i)]\delta(x_i-x)-\frac{M'}{(b_1-a_1)}E$$

where E is the statistical error due to Monte-Carlo method and E is also dependent on x. The interpolant f_A is defined by (Eq. 10). So, the absolute error between f(x) and f_A(x) is,

| f(x)- f_A(x) | =

$$\left|\frac{\sum_{i=1}^{M'}[f(x_i)]\delta(x_i-x)-\frac{M'}{(b_1-a_1)}E}{\sum_{i=1}^{M'}\delta(x_i-x)}-\frac{\sum_{i=1}^{M'}[f(x_i)]\delta(x_i-x)}{\sum_{i=1}^{M'}\delta(x_i-x)}\right|$$

$$= \left| \frac{E}{\frac{(b_1-a_1)}{M'}\sum_{i=1}^{M'}\delta(x_i-x)} \right| = \frac{|E|}{\frac{(b_1-a_1)}{M'}\sum_{i=1}^{M'}\delta(x_i-x)}$$

Note that the denominator is always "positive" and "not equal to zero", and for 2-dimension case, the above equation is generalized to,

$$|f(x_1, x_2)-f_A(x_1, x_2)| = \frac{|E|}{\frac{(b_1-a_1)(b_2-a_2)}{M'}\sum_{i=1}^{M'}\delta(x_{1i}-x_1)\delta(x_{2i}-x_2)}$$

(Eq. 16)

where $\delta(x_{1i}-x_1)$ and $\delta(x_{2i}-x_2)$ are defined by (Eq. 3).Note that on the right hand side of (Eq. 16), the denominator can be computed for the requested location (x_1, x_2) and the numerator E is governed by the Central Limit Theorem. As said earlier about (Eq. 6), the deviation error E will exceed $\pm\frac{\sigma}{\sqrt{N}}$ with probability 31%, $\pm\frac{2\sigma}{\sqrt{N}}$ with probability 4.5%, and $\pm\frac{3\sigma}{\sqrt{N}}$ with probability 0.3%. Again, N is equal to the input location number M '. All needs to be done is to find the value of σ and we shall do so as follows.

Let us recall Eq. 5,

$$\overline{Z} = \int_{-\infty}^{+\infty} Z(x')g(x')dx'$$

Compare the above integral with Eq. 7,

$$0 = \int_{a_1}^{b_1} [f(x') - f(x)]\,\delta(x'-x)dx'$$

We obtain,

$$\overline{Z} = 0$$

$$(b_1-a_1)\ [f(x') - f(x)]\delta(x'-x) = Z(x')$$

and the density function,

$$g(x') = \begin{cases} 1/(b_1-a_1), & \text{for } a_1 \le x' \le b_1 \\ 0, & \text{otherwise} \end{cases}$$

By definition $\sigma^2 = \overline{(Z - \overline{Z})^2}$ and $\overline{Z} = 0$, it gives $\sigma^2 = \overline{Z^2}$.

$$\sigma^2 = \overline{Z^2} = \int_{-\infty}^{+\infty} (Z(x'))^2 g(x') dx' \qquad \text{(Eq. 17)}$$

(Eq. 17) can not be calculated because it involves the unknown function $f(x^{'})$. However, the sample σ^2 can be computed by use of,

$$\sigma^2 = \overline{Z^2} = \frac{1}{M'} \sum_{i=1}^{M'} Z_i^2 = \frac{1}{M'} \sum_{i=1}^{M'} [(b_1 - a_1)[f(x_i) - f(x)]\delta(x_i - x)]^2$$

(Eq. 18)

The 2-dimensional case of (Eq. 18) has the following form,

$$\sigma^2 = \overline{Z^2} = \frac{1}{M'} \sum_{i=1}^{M'} Z_i^2 =$$

$$\frac{1}{M'} \sum_{i=1}^{M'} \{(b_1 - a_1)(b_2 - a_2)[f(x_{1i}, x_{2i}) - f(x_1, x_2)]\delta(x_{1i} - x_1)\delta(x_{2i} - x_2)\}^2$$

(Eq. 19)

The above equation and (Eq. 16) were used to calculate the uncertainty for the interpolation output results of SIC 2004 exercise. The function value $f(x_{1,} x_2)$ was set to the measured, true value which was provided by SIC 2004 (1st_file_true_values.csv & 2nd_file_true_values.csv). In the event when no measured value is given, then $f(x_{1,} x_2)$ can be set to the interpolated value, $f_A(x_1, x_2)$. We note in passing that, by reducing the delta width Δ_1 and Δ_2 used in (Eq. 13) towards zero, the interpolation semi-norm, $\sum_{i=1}^{M'} \left| f(x_{1i}, x_{2i}) - f_A(x_{1i}, x_{2i}) \right|$ approaches zero as well. This property of semi-norm approaching zero will remain true in our formulation for any higher dimensional space.

2.1 *Use of prior information*

Due to the simplicity of DMC interpolation process, the prior information (10 days of measurements) were not used at all, except for tuning the delta width explained in the following section. DMC method treats different input data sets the same manner, independent of anomaly which may exist in the data set.

2.2 *Tuning the algorithms*

The only thing which needs to be tuned in using DMC interpolant is the "delta width" value. We present below how Δ_1 and Δ_2 values are estimated for SIC 2004 exercise.

By using (Eq. 15), and $(b_1 - a_1)$ for x-axis is set approximately at 360,000 meters and $(b_2 - a_2)$ for y-axis at 700,000 meters for SIC 2004 exercise. It is straightforward to calculate and find Δ_1 and Δ_2 values. They are,

$$\Delta_1 = \Delta_2 = 11{,}300 \text{ meters}$$

These are first-cut values for Δ_1 and Δ_2. Now, we can fine tune these values by firstly finding the maximum peak and minimum valley locations of prior radioactivity measurement information of any day (Due to the shortage of time, only one day measurement was used. We "randomly" chose the second day measurement data.), secondly finding the requested locations that are closest to the maximum peak and minimum valley locations, thirdly computing repetitively the interpolated value at these requested locations by use of (Eq. 13) by gradually changing and reducing Δ_1 and Δ_2 values, and lastly stopping the previous step when the interpolated value is "reasonably" close to the data value at the peak and the valley location. To decide what value is considered as "reasonable" largely depends on the practitioner's experience and judgment for the problem. It should be said that to choose smaller values of Δ_1 and Δ_2 than necessary will end up producing interpolation results with large variations (That is, not smooth) and large σ value across the supported domain. Generally speaking, smaller delta width will produce larger range of interpolated function values and bigger delta width will produce smaller range of interpolated function values. At the end of the above tuning process, we chose $\Delta_1 = \Delta_2 = 4000$ meters for the exercise.

3. RESULTS

In the following analysis, two input data files were used in the exercise. The second data file is the "joker" data set. Please note that only 800 (not 808) estimated values were generated in our work***. The last 8 locations in the output location file were not used.***

3.1. *Overall results*

The table below gives the minimum, maximum, mean, median and standard deviation of the 800 estimated values and the observed values for both data set 1 and 2.

N = 800 (not 808)	**Min.**	**Max.**	**mean**	**median**	**std. dev.**
Observed (first data set)	57	180	98.21	99	19.97
Estimates (first data set)	66.49	145.53	96.91	99.57	14.34
Observed (second data set)	57	1528.2	105.6	99	84.01
Estimates (second data set)	67.15	775	108.95	102.14	59.5

Table 1. *Comparison of the estimated and measured values (nSv/h).*

The mean absolute error (MAE), the bias (or mean error ME), and the root mean squared error (RMSE) of the predictions at the *n* = 800 locations are given in the table below. These quantities are defined as,

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|f_i^* - f_i\right|,$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}\left(f_i^* - f_i\right),$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(f_i^* - f_i)^2},$$

where f^*_{i} (defined as the estimated value at location i and where f_{i} is the true value. Note that earlier f^*_{i} is defined as $f_A\ (x_{1i}\ ,\ x_{2i}\ )$ and f_i is defined as $f\ (x_{1i}\ ,\ x_{2i}\ )$. Pearson's r coefficient of correlation between the estimated and true values is also given.

Data sets: (N = 800)	**MAE**	**ME**	**Pearson's *r***	**RMSE**
First data set	9.67	-1.29	0.75	13.21
Second data set	19.91	3.26	0.61	66.80

Table 2. *Comparison of the errors.*

1) Estimations:

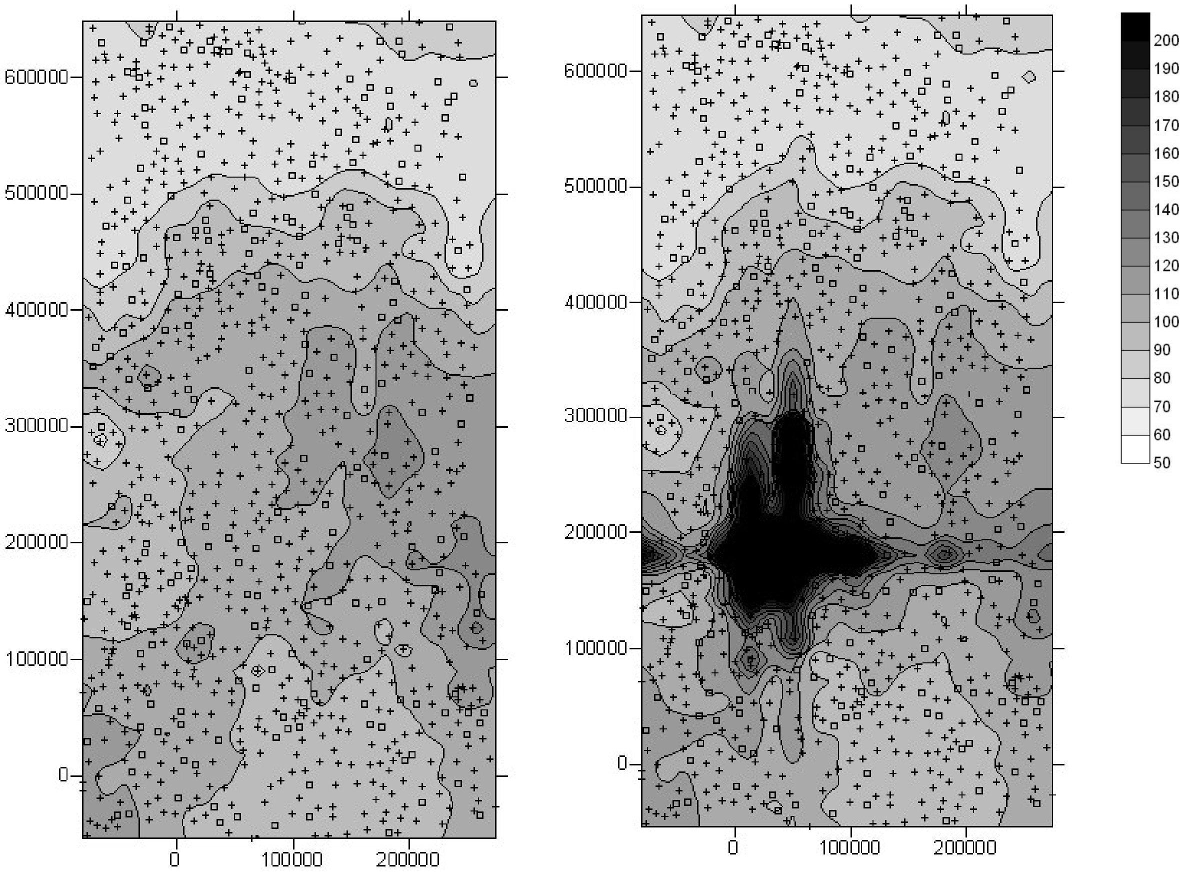


Figure 1. *Isoline levels (nSv/h) for the 1st set (let) and the 2nd set (right).*

2 maps above present contour lines obtained for the two datasets. The colour scale is black and white, with colour white for level 50 and black for levels above 200. Intervals are of 10 nSv/h. Crosses point to the locations of the estimated values; empty squares are used to indicate the locations of the input values.

The maps in Figure 1 are generated by use of the estimated values at 800 random locations. With the use of DMC interpolant ($\Delta_1 = \Delta_2 = 4000$ meters) as “gridding method”, the interpolated values for 800 grid nodes covering the entire area of study are made. The grid has 20 columns and 40 rows. Then, the contour lines are generated based on the values of 800 grid nodes.

2) Uncertainty:

Two maps are presented below showing the levels of uncertainty that are associated to the estimations described in the previous figures which display contour lines obtained for the two datasets. The uncertainty value $|f(x_{1,}\ x_2)\text{-}\ f_A(x_1,\ x_2)|$ is defined by (Eq. 16) where E value is set to the absolute value of $\pm\frac{2\sigma}{\sqrt{N}}$. Thus, the confidence level associated with the uncertainty value is 95.5%. Note that (Eq. 19) was used to calculate σ for 800 output random locations.

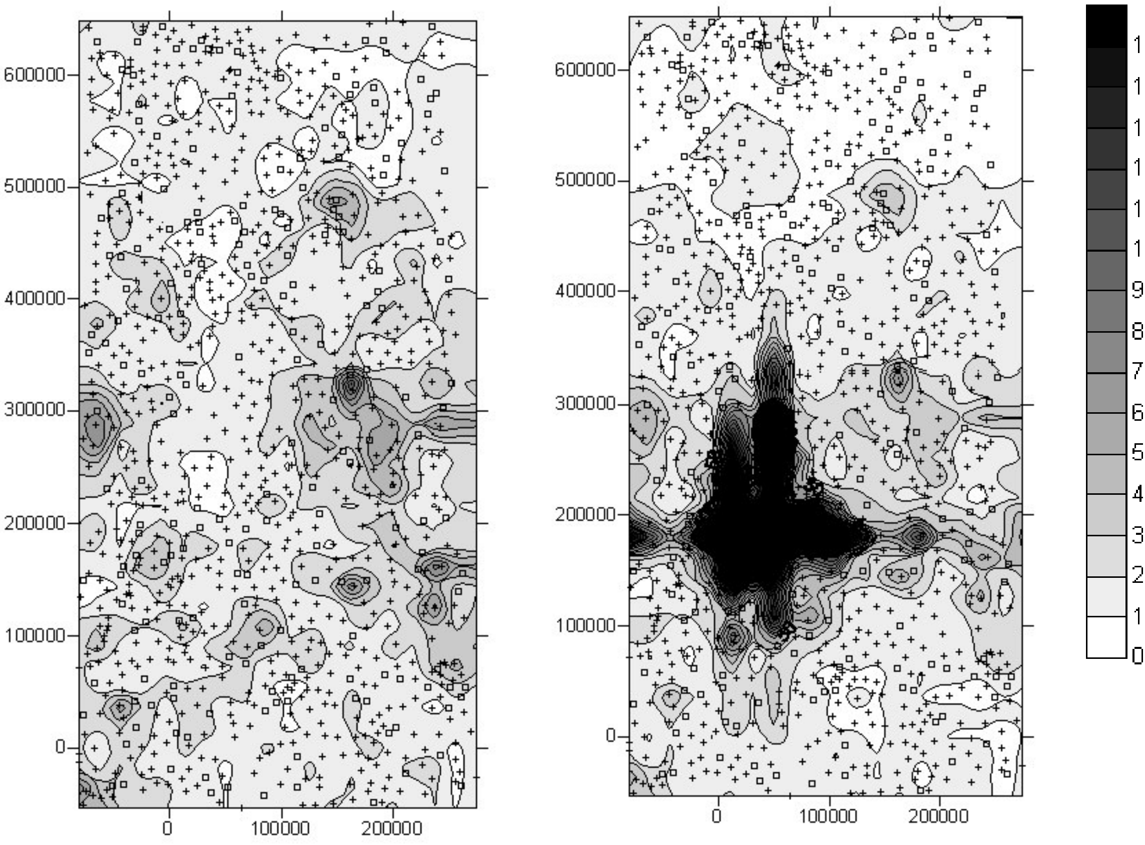


Figure 2. *Isoline levels showing the uncertainty (95.5% confidence level) associated to the estimations obtained for the 1st set (let) and the 2nd set (right).*

For the first estimation data set, the uncertainty value has maximum value 92.181 at location # 858, and minimum value 0.819 at location # 671 among all estimation locations. The median value is 9.771 . For the second estimation data (joker) set, the uncertainty value has maximum value 1364.648 at location # 545, and minimum value 1.211 at location # 840 among all estimation locations. The median value is 12.058 .

3.2. *Detecting anomalies and outliers.*

The following figure displays the estimates obtained for the 2nd set in 3 D, $\Delta_1 = \Delta_2$ = 4000 meters were used.

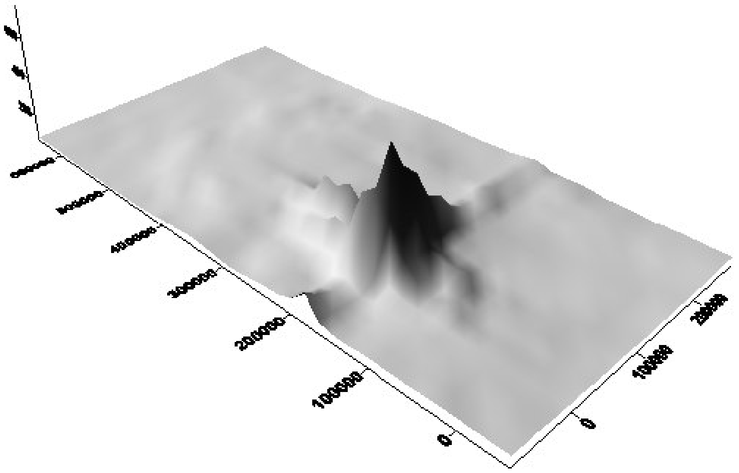


Figure 3 *3D map showing extreme values found in the 2nd set (vertical scale in nSv/h)*

4. DISCUSSION

Dirac-Monte Carlo method is a simple, though guided by the higher-level mathematics, interpolation method to use. The interpolant employed in DMC method is defined by (Eq. 12), and with the choice of rational function (Eq. 3) for the representation of Dirac delta function, (Eq. 13) is established and it can be easily programmed for computation. Once the number of input location is given, the user only needs to preset Δ_1 and Δ_2 values for completion of automating (Eq. 13).

As can be seen in Table 1, DMC produces the interpolated values which were constrained between the maximum and the minimum values of the given, measured input data. DMC method, similar to Kriging, is capable of producing uncertainty and confidence level data whereas most of other interpolation methods can not do so. Moreover, It is much faster and more direct to use DMC to calculate uncertainty values than Kriging.

One thing needs to be said about Figure 1 pertaining to the 2nd (joker) input data set. It can be seen that on the left edge near 20000 on the Y-axis, there is an artifact "dark blob" (Note that the same artifact shows up in Figure 3). In addition, there are 2 cross-shape, rather than circular-shape, dark blobs. This is due to the fact that the joker data set has two high input values (location 339, with measured value1499 and location 549 with value 1070.4). Also, it is because that the DMC interpolant is not "distance" dependent but rather "coordinate separation" dependent, as emphasized before. Whenever an extreme value (positive or negative) exists at an input location (x, y), this particular input will give long range influence along the horizontal and vertical direction on the "cross" centered at (x, y). In order to curb this long range effect, it is essential to enlarge Δ_1 , Δ_2 values accordingly. For example, in the current situation, the delta width values used, ($\Delta_1 = \Delta_2$ = 4000 meters), need to be increased. If we use $\Delta_1 = \Delta_2$ = 8000 meters, the artifact feature is reduced. (see Figure 4 below)

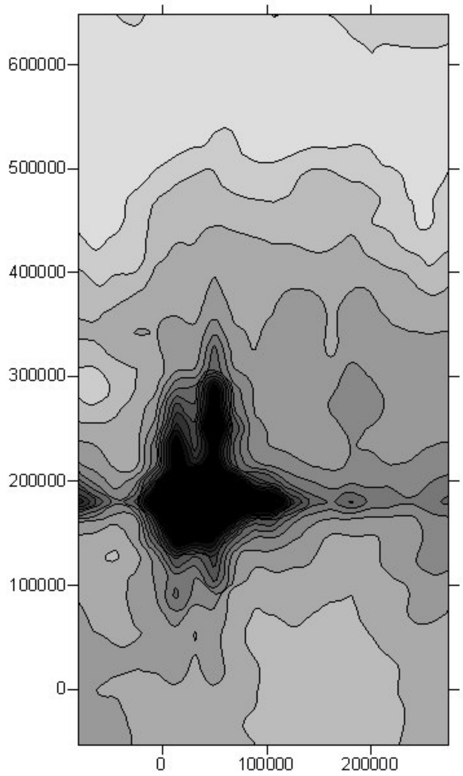


Figure 4 (Color scale is the same as Figure 1)

The total time for our conducting the exercise between download input data from and upload output data to SIC 2004 web site is about 10 minutes. Thus, it roughly takes 5 minutes for each input data set to be processed. Within the 5 minutes time, the computer operator needs to connect to the Internet, perform input/output data file transfer, zip input/output file, link the input/output files to the software program on the local PC, execute the interpolation software program, and send back output files. The next paragraph describes the average time pertaining to the execution of the interpolation software program.

The average computational time for both input data sets is about “5 to 6 seconds”. Just about all of this time was used for reading input data file and formatting the output data file within the software. The actual computing time for “interpolation”, that is, executing (Eq. 13), is much less than a fraction of one second. All computations were performed on a standard PC with about 2 GHZ CPU. One can clearly see that DMC method can be used for real-time computational applications, ideal for handling emergency situations. For more input locations or higher dimensionality, this computing-time advantage will become even more pronounced as compared with any other interpolation methods.

5. CONCLUSIONS

DMC interpolation method is a new method. It has been employed to carry out SIC 2004 exercise to generate interpolated data values as well as the associated measures of merit (such as **MAE, ME, Pearson’s** *r,* **RMSE** *).* Furthermore, detailed uncertainty findings for 800 output locations have been given for the confidence level 95.5% in a straightforward and candid manner by adhering to the formulation of DMC method.

We highlight ongoing efforts and possible development for DMC method below.

(1) Δ_1 and Δ_2 values could and should be preset differently for some applications. In addition, they do not have to be constant and could depend upon the location where the interpolated value is computed. Note that delta width values directly control the long-range effect mentioned before. This is one area where more study will be carried out.

(2) Employ DMC interpolants for applications by means of polar, spherical and cylindrical coordinates (ref. 5). In radar weather prediction analysis, the so-called "adaptive Barnes interpolant" was expressed in terms of spherical coordinates. But its mathematical form was not treated analytically. (ref. 12) With DMC interpolant expressed in terms of spherical coordinates, the practitioner can use it with ease for global weather/environment modelling and prediction.

(3) The input locations where measurements were taken, though random, can be preset by use of "quasi-Monte Carlo" sampling (ref. 13). By doing so, the interpolated values will be more accurate.

(4) Establish subscription center at RDIC to offer DMC capabilities online on the web to the commercial market.

Acknowledgments

Thanks to AI-GEOSTATS for the opportunity to join SIC 2004 exercise. Critical comments and suggestions on Dirac-Monte Carlo method are most welcome. Figures (1-4) are generated by use of Surfer program (courtesy of Golden Software, Colorado, USA).

Codes

The primary objective of this paper is to present Dirac-Monte Carlo interpolation method to the geostatistics community. At the same time, we promote the concept of performing real-time computation on the web by means of standard web pages (HTML and Javascript) as demonstrated at RDIC (http://www.fanginc.com/main.htm). Numerical interpolation software used in the exercise was programmed in terms of web pages. Input and output files required to run the software had been linked to the web pages by simply using "copy" and "paste" PC WINDOWS features. Inquiry regarding our software can be sent to the email address "fanginc@gte.net".

REFERENCES

1. Random Data Interpolation Center (RDIC), *Introduction page*, http://www.fanginc.com/introduction.htm, 2002.
2. DMTI - White paper, *"Interpolation Methods for Generalizing Ozone Concentration Patterns*", http://www.dmtispatial.com, Ontario, Canada.
3. R.J. Renka, "*Multivariate interpolation of large sets of scattered data"*, ACM Trans. On Mathematical Software, 14:2, June 1988, pp 139-148
4. Wand, M. P. & Jones, M. C. (1995). Kernel Smoothing, Vol. 60 of Monographs on Statistics and Applied Probability, Chapman and Hall, London
5. FANG, INC., *"Interpolation Model (Dirac-Monte Carlo Method)",* SBIR proposal to

National Science Foundation, USA 2004 - Technical content of this proposal shall be made available at RDIC in the near future.

6. FANG, INC., “Supplementary Material for Random Data Interpolation”, http://www.fanginc.com/texas2.doc, 2004.
7. P. Dirac, The Principles of Quantum Mechanics, Oxford University Press, 1930.
8. S.J. Yakowitz, Computational Probability and Simulation, Addison- Wesley, 1977.
9. D. Gillespie, The Monte Carlo Method of Evaluating Integrals, ADA-005891, DTIC, USA, 1975.
10. Nadaraya, E. A. On estimating regression. Theory of Probability and Its Applications, Vol. 9, 1964, pp. 141-142.
11. Wand, M. P. & Jones, M. C. (1995). Kernel Smoothing, Vol. 60 of Monographs on Statistics and Applied Probability, Chapman and Hall, London
12. *Quantitative Precipitation Estimation and Segregation Using Multiple Sensors*, http://www.norman.noaa.gov/publicaffairs/backgrounders/backgrounder_qpe.html
13. Niederreiter, H. & P. Hellekalek & G. Larcher & P. Zinterhof, Eds. (1998): "*Monte Carlo and Quasi-Monte Carlo Methods 1996*", Springer-Verlag New York, Lectures Notes in Statistics, 1998, 448 pp.