Advances in Geosciences 地球科学前沿, 2018, 8(6), 1051-1058
Published Online October 2018 in Hans. http://www.hanspub.org/journal/ag
https://doi.org/10.12677/ag.2018.86115

Hans汉斯

The Correction of String Effect by Finite Domain Kriging

Ziqi Li[1], Yongbiao Wang[1,2], Shaohua Li[1*], Yuangang Liu[1], Peiqing Lian[3]

[1]School of Geoscience, Yangtze University, Wuhan Hubei
[2]Research Institute of Zhanjiang Company, CNOOC, Zhanjiang Guangdong
[3]Petroleum Exploration and Production Research Institute, SINOPEC, Beijing
Email: 318534814@qq.com, *534354156@qq.com

Received: Oct. 2nd, 2018; accepted: Oct. 18th, 2018; published: Oct. 25th, 2018

Abstract

In the process of kriging estimation with finite strings of data, it can be frequently observed that outlying data in the strings receive higher weights than all other data. This counterintuitive weighting is referred as string effect. Taking simple kriging as an example, we adopt the Finite Domain Simple Kriging method (FDSK) to correct the string effect of simple kriging algorithm, and compared with the Distance Constrained Kriging method (DCSK). The FDSK estimators are linear combinations of several traditional kriging estimators. The kriging weights to be used for FDSK are the average of the weights from the traditional kriging runs. FDSK approach has been programmed by FORTRAN language and tested with a small example. The new method can correct the string effect of the traditional kriging method with a high estimate accuracy which has important realistic significance to improve the accuracy of the estimate in the mining and petroleum industries.

Keywords

String Effect, Simple Kriging, Finite Domain Kriging, Distance Constrained Kriging

采用有限域克里金方法进行校正条带效应

李紫琪[1]，王勇标[1,2]，李少华[1*]，刘远刚[1]，廉培庆[3]

[1]长江大学地球科学学院，湖北 武汉
[2]中海石油(中国)湛江分公司，广东 湛江
[3]中国石化石油勘探开发研究院，北京
Email: 318534814@qq.com, *534354156@qq.com

收稿日期：2018年10月2日；录用日期：2018年10月18日；发布日期：2018年10月25日

*通讯作者。

文章引用: 李紫琪, 王勇标, 李少华, 刘远刚, 廉培庆. 采用有限域克里金方法进行校正条带效应[J]. 地球科学前沿, 2018, 8(6): 1051-1058. DOI: 10.12677/ag.2018.86115

摘　要

在克里金估值的过程中，在条件数据出现带状分布的情况时，两个端点条件数据的权值容易呈现偏大的条带效应。本文以简单克里金(SK)为例，针对其中产生的条带效应，运用有限域克里金方法(FDSK)对其进行了校正，并将距离约束克里金方法(DCSK)与该方法进行对比。FDSK方法的估计量是多个传统克里金估计量的线性组合，FDSK估计量中用到的权值是这些传统克里金权值的平均值。将FORTRAN的编程语言运用于FDSK方法，并利用FDSK方法进行了实例数值模拟。结果表明新方法不仅消除了传统克里金方法引起的条带效应，而且具有较高的估计精度，优于距离约束克里金方法。该方法的应用能够提高矿业和石油行业中估值和模拟计算的精度。

关键词

条带效应，简单克里金，有限域克里金，距离约束克里金

1. 引言

克里金插值法名字的由来是以南非矿业工程师 D. G. Krige 的名字命名，其为最优线性内插法[1] [2]。从统计学角度来讲，克里金插值方法是基于变量相关性和变异性，以无偏、最优估计为特点，在有限区域范围内对区域化变量进行估值；在插值方面上来看，克里金插值方法是以无偏内插估计、求线性最优为特点，对空间分布的数据进行估值。在地质行业中，一般以钻井得到所需的主要条件数据，获得的条件数据通常出现条带状分布的情况，该原因导致在应用克里金插值的过程中会出现条带效应的现象，即表现为在有限的呈条带分布数据系列的两端，其数据会呈现出偏大的权值[3] [4]。许多学者在该方面做了大量的研究，如肖克炎、张晓华、王全明等利用改进的克里格方法分离重力区域异常与局部异常；高美娟、朱庆忠、张淑华等利用贝叶斯—克里金估计技术进行储层参数预测；汪保、孙秦对克里金模型的可靠度进行了改进。克里金插值方法不仅在多个行业的数据成图方面广受欢迎[5] [6] [7]，而且在以变差函数为基础的系列随机模拟算法方面用来构建待估点处局部条件概率的分布，为储层随机建模技术的重要组成部分[8] [9] [10]。因此为了使克里金估值变得更为精准，需要消除条带效应。本文以简单克里金为例，通过有限域克里金方法对克里金估值得到的权值进行校正，对实例数据进行数值模拟验证，并将距离约束克里金方法与其进行了比较。

2. 简单克里金原理[11]

简单克里金法即在区域化变量 $Z(x)$ 的数学期望 $E\left[Z(x)\right]$ 为常数且已知的情况下建立的克里金法。由于 $E\left[Z(x)\right]=E\left[Z(x+h)\right]=m$ 已知，令 $Y(x)=Z(x)-m$，则 $E\left[Y(x)\right]=E\left[Z(x)-m\right]=0$，其协方差 $E\left[Y(x),Y(y)\right]=Cov(x,y)$，那么对 $Z(x_0)$ 的估计可转化为对 $Y(x_0)$ 的估计，只要求出 $Y^*(x_0)$ 就可以得到 $Z^*(x_0)$。

$$Y^*(x_0)=\sum_{i=1}^{n}\lambda_i Y(x_i)$$

式中：$Y^*(x_0)$是无偏估计量，λ_i是权重，$Y(x_i)$是误差。

估计方差

$$\sigma_E^2 = E\left[Y(x_0) - Y^*(x_0)\right]^2$$

式中：σ_E是估计方差，E是数学期望。

为了使达到最小值，按求极值原理，对求偏导数。即：

$$\frac{\partial \sigma_E^2}{\partial \lambda_i} = -2Cov(x_i, x_0) + 2\sum_{j=1}^{n} \lambda_j Cov(x_i, x_j) = 0$$

式中：$Cov(x_i, x_0)$是数据点 x_i 与数据点 x_0 之间的协方差，$Cov(x_i, x_j)$数据点 x_i 与数据点 x_j 之间的协方差。

进而得到简单克里金方程组：

$$\sum_{j=1}^{n} \lambda_j Cov(x_i, x_j) = Cov(x_i, x_0)$$

求解，即可得到简单克里金权系数，同时可得到简单克里金估计方差：

$$\sigma_E^2 = Cov(x_0, x_0) - \sum_{i=1}^{n} \lambda_i Cov(x_i, x_j)$$

3. 简单克里金的改进

3.1. 条带效应

在地质行业中，通常是通过钻孔或钻井来获得主要的条件数据，这些条件数据通常是条带状分布的情况。在该情况下，运用克里金对条件数据进行插值，如果被插值的数据为有限条带数据时，则会出现条带效应[3] [4]：在条带数据中，两端距离估计点最远的样品点获得的权值偏大，反而离估计点较近距离的权值较小。而该情况与实际情况是不相符的，因此有必要对条带效应进行校正。本文采用有限域克里金方法来校正这种条带效应，对该方法进行了实例数值模拟验证，并与距离约束克里金方法进行了比较。

3.2. 有限域克里金方法(FDSK)原理

加拿大 Olena Babak 博士对条带效应进行校正的方法进行了探讨，总结出针对条带效应可利用有限域克里金方法对其进行校正，并且给出了计算例子[4]。本文在其基础上，编程实现了三维空间中条带效应的校正，并进行了实例验证。

假设在一个条带数据中有 n 个位于相邻位置 $u_i\ (i = 1, 2, \cdots, n)$ 处排成一列的数据点，对待估点 u_0 处未取样样本 Z 的变量值采用有限域克里金方法进行估值，该方法的基本原理如下所示：

对条带数据，在待估点处的 FDSK 估计量是：

$$Z_{FDSK}^*(u_0) = \frac{1}{n}\sum_{k=1}^{n}\left[\lambda_{sk,i}^k(u_0) Z_i^k(u_0) + \left(1 - \sum_{i=1}^{k} \lambda_{sk,i}^k(u_0)\right) m\right]$$

式中 $Z^k = (Z_1^k, Z_2^k, \cdots, Z_k^k), k = 1, \cdots, n$，表示在条带数据中 k 个最接近待估点 u_0 的数据，Z^k 中的元素经过了排序，Z_1^k 是最接近待估点的样本数据，Z_2^k 是次接近待估点的样本数据，以此类推；

$\lambda_{SK}^{K,\mathrm{T}} = (\lambda_{SK,1}^K, \cdots, \lambda_{SK,k}^K)^{\mathrm{T}}, k = 1, \cdots, n$，表明采用简单克里金方程组对待估点 u_0 进行估值得到的简单克里金权值：

$$\sum_{i=1}^{k} \lambda_{sk,i}^k Cov(Z_j^k, Z_i^k) = Cov(Z_j^k, Z(u_0)), j = 1, \cdots, k$$

式中：$Cov\left(Z_j^k, Z_i^k\right)$是样本数据与样本数据的协方差，$Cov\left(Z_j^k, Z\left(u_0\right)\right)$是样本数据与待估点的协方差，并且，$i, j=1,\cdots,k$，$k=1,\cdots,n$。

有限域简单克里金的估计量$Z_{FDSK}^*\left(u_0\right)$可以简化为：

$$Z_{FDSK}^*\left(u_0\right)=\sum_{i=1}^{n}\lambda_{FDSK,i}\left(u_0\right)+\left(1-\sum_{i=1}^{n}\lambda_{FDSK,i}\left(u_0\right)\right)m$$

式中：$\lambda_{FDSK,i}^{\mathrm{T}}\left(u_0\right)=\left(\lambda_{FDSK,1}\left(u_0\right),\cdots,\lambda_{FDSK,n}\left(u_0\right)\right)^{\mathrm{T}}=\left(\lambda_{FDSK,1},\cdots,\lambda_{FDSK,n}\right)^{\mathrm{T}}$

是有限域简单克里金权值，由下面的公式计算得到：

$$\lambda_{FDSK,i}=\frac{1}{n}\sum_{k=1}^{n}\tilde{\lambda}_{SK,i}^{k}$$

有限域简单克里金估计方差如下：

$$\sigma_{FDSK}^2\left(u_0\right)=\sigma^2-2\sum_{i=1}^{n}\lambda_{FDSK,i}Cov\left(Z\left(u_i\right),Z\left(u_0\right)\right)+\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_{FDSK,i}\lambda_{FDSK,j}Cov\left(Z\left(u_i\right),Z\left(u_0\right)\right)$$

3.3. FDSK 算法实现

FDSK 方法实现的具体步骤如下：

1) 计算已知点与待估点之间的一组距离$d_1,d_2,\cdots,d_n$，并按升序将这些距离进行排列，令 m = (排序后的距离的脚标)；

2) 将已知点按距离进行排列$Z_1,Z_2,\cdots,Z_n$；其中，离待估点距离最近的样本点为 Z_1，离待估点距离次近的样本点为 Z_2，……，表示离待估点距离最远的样本点为 Z_n；

3) 利用简单克里金方法(SK)求出$k\left(1,2,\cdots,n\right)$个与待估点最接近的已知点的权值，$\lambda_{SK}^k=\left(\lambda_{SK,1}^k,\cdots,\lambda_{SK,i}^k\right), i=1,\cdots,k$ 并且当$i>k$时，令$\lambda_{SK,i}^k=0$；

4) 令$\lambda_F=\left(\lambda_{F,1},\lambda_{F,2},\cdots,\lambda_{F,n}\right)$，

$$\lambda_{F,i}=\frac{1}{n}\sum_{k=1}^{n}\lambda_{SK,m(i)}^{k}$$

式中：$m(i)$是步骤 1)中 m 的第 i 个元素；

5) 则λ_F中的元就是我们所需要的 FDSK 权值。

4. 权值计算检验

通过一个简单的例子，对 FDSK 的计算效果进行验证。假定有 7 个样品点呈条带状分布，位置分别为(1，0)，(2，0)，(3，0)，(4，0)，(5，0)，(6，0)，(7，0)，对位于(4，20)的点进行估值，在估值的过程中，采用基台值为 1、变程为 40 的球状变差函数模型，得到 SK、DCSK 和 FDSK 的权值及对应的估计方差如图 1。从图中可以看到：1) SK 得到的权值具有明显的两端权值大中间小的条带效应，而 DCSK 和 FDSK 得到的权值消除了这种条带效应；2) DCSK 和 FDSK 的估计误差都与 SK 的估计误差很接近；3) 由于 DCSK 受到距离的约束导致有不平滑的权值产生，而 FDSK 的权值是多个 SK 权值的平均值，故每次估计都是最优的，不会产生不平滑的权值。因此 FDSK 方法不仅在消除 SK 条带效应方面具有良好的效果，在估计精度方面也具有较高的可靠信，而且每次估值均为最优，优于 DCSK。

图 2 为将条件数据个数改变的情况下，对待估点(位于中间)采用 FDSK 进行估值，进而获得权值的分布结果。从权值的分布结果可以得出，FDSK 具有良好的稳定性，在改变条件数据个数的情况下，依然能保持权值分配的趋势一致，对两端条件数据的权值均能起到修正的作用。在未位于中间位置待估点

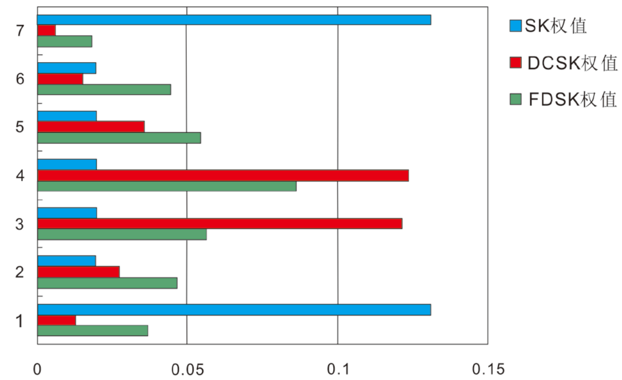


Figure 1. Weights with a nugget effect of 0
图 1. 块金效应为 0 时的权值结果图

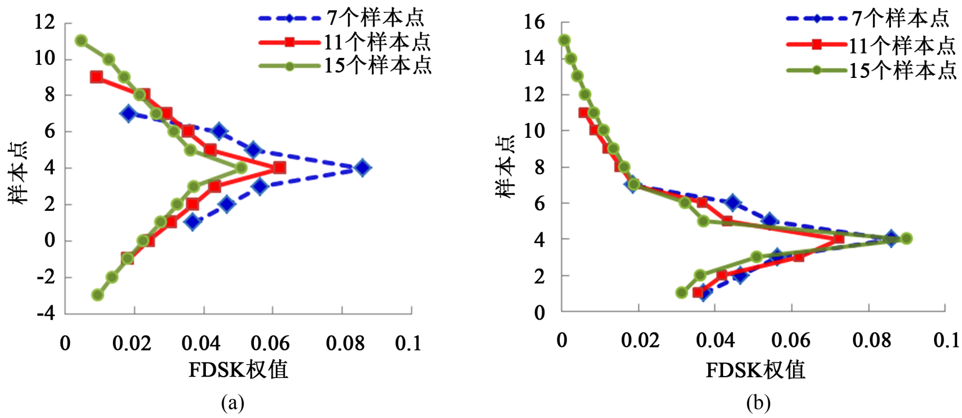


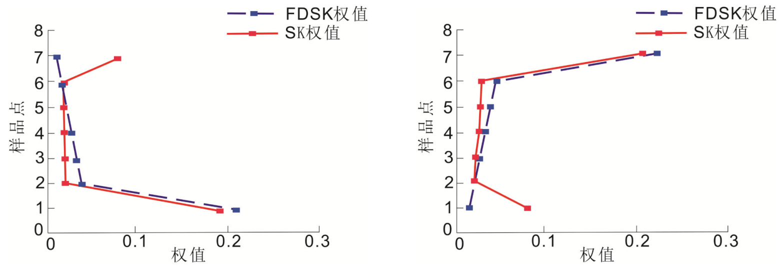
Figure 2. Weights with different numbers of Strings of data. (a) The number of conditional data increases to both ends; (b) The number of conditional data increases toward one end
图 2. 具有不同权重条件数据的权值。(a) 条件数据个数向两端增加；(b) 条件数据个数向一端增加

的权值分配方面，FDSK 也能对其进行较好的校正。如图 3(a)与图 3(b)，待估点的位置平行于条件数据的某一端点，FDSK 便对另外一个端点条件数据的权值进行了修正。

5. 数值模拟计算检验

假设有一组数据呈条带状分布(如图 4)，采用 DCSK、FDSK 和 SK 方法对其分别进行数值模拟计算，得到的结果如图 5。在估值过程中，采用块金效应为 0，基台值为 1、变程为 40 的球状变差函数模型；网格划分为 15 × 15。

对比图 5(a)~(c)可以看到：在中间偏上和偏下的位置，DCSK 方法和 FDSK 方法比 SK 模拟得到的数值偏大，这是因为 DCSK、FDSK 方法校正了 SK 的条带效应。图 5(d)和图 5(e)分别是 DCSK、FDSK 与 SK 模拟结果的差异图(即条带效应校正的结果图)。比较图 5(d)和图 5(e)两个图可以清楚的看到高值和低值的分布区域，这些区域表明这些地方受到克里金的条带效应的影响，从图 5(d)和图 5(e)中可以得出 DCSK 和 FDSK 这两种方法都是通过局部去校正。这两个图的相同之处是：离条带数据越远的区域权值的修正效果越大，这是因为待估点离条件数据越远条带效应越大；当待估点非常接近条带数据时，传统克里金方法基本上不会出现条带效应的情况。这两个图的差异之处是：DCSK 对位于中上、中下偏两边的区域的校正结果比较明显，而 FDSK 对中间偏上、下区域的校正结果比较明显。其原因在于 DCSK 受到距离的约束导致次优的估计引起了不平滑的权值，而 FDSK 没有这种不平滑的权值。



(a) (b)

Figure 3. Weights in different locations of the estimation point. (a) An estimate of the point to be estimated at (1, 20); (b) An estimate of the point to be estimated at (7, 20)
图 3. 待估点不同位置时的权值。(a) 对(1，20)处的待估点的估计；(b) 对(7，20)处的待估点的估计

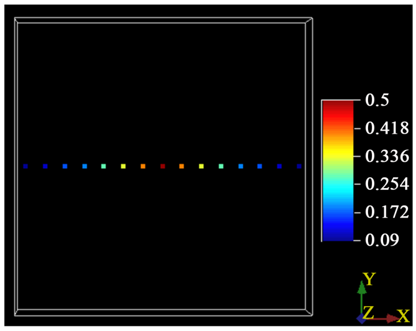
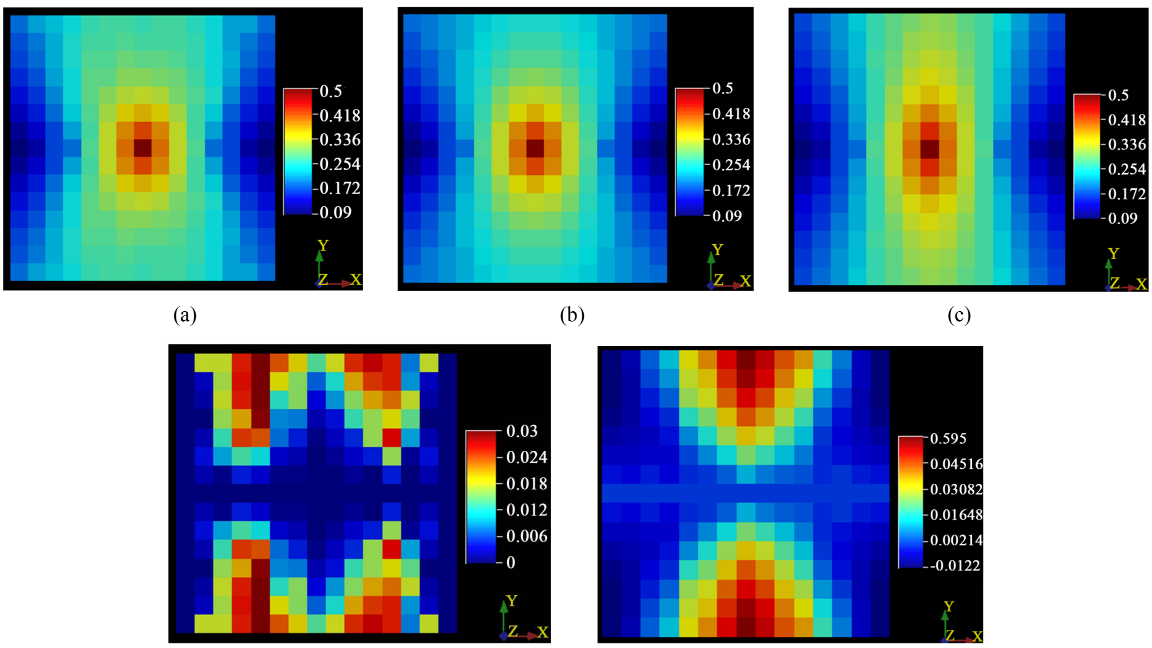


Figure 4. Strings of data
图 4. 条带数据



(a) (b) (c)

(d) (e)

Figure 5. Simulation results. (a) Results of DCSK simulation; (b) Results of SK simulation; (c) Results of FDSK simulation; (d) Difference map of simulation results between DCSK and SK; (e) Difference map of simulation results between FDSK and SK
图 5. 数值模拟结果图。(a) DCSK 模拟的结果；(b) SK 模拟的结果；(c) FDSK 模拟的结果；(d) DCSK 与 SK 模拟结果的差异图；(e) FDSK 与 SK 模拟结果的差异图

6. 实测数据应用

从某油田提取 4 口井的共 60 个孔隙度数据，为了便于模拟计算对实际数据进行处理，处理后的数据分布如图 6 所示。分别用 DCSK、FDSK 和 SK 方法对处理后的数据进行模拟(在估值过程中采用块金效应为 0，变程为 100，基台值为 1 的球状变差函数模型；网格划分为 15 × 15 × 15)，图 7 为模拟结果，其中 FDSK 与 SK 的模拟结果差异图分别为图 7(a)和图 7(b)，图 7(c)和图 7(d)分别是图 7(a)和图 7(b)在 z 方向上的一个切片图。从图 7(c)和图 7(d)可以清楚的看到高值和低值的分布区域，表明这些区域受到克里金条带效应的影响，并且这两种方法都对条带效应进行了有效的校正；比较图 7(c)和图 7(d)可以发现：图 7(d)比图 7(c)中的高值和低值区域的范围大，说明 FDSK 比 DCSK 的校正范围大，即 FDSK 比 DCSK 有更强的校正效果。

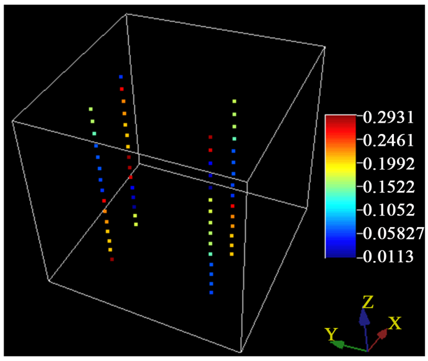


Figure 6. Porosity data
图 6. 孔隙度数据

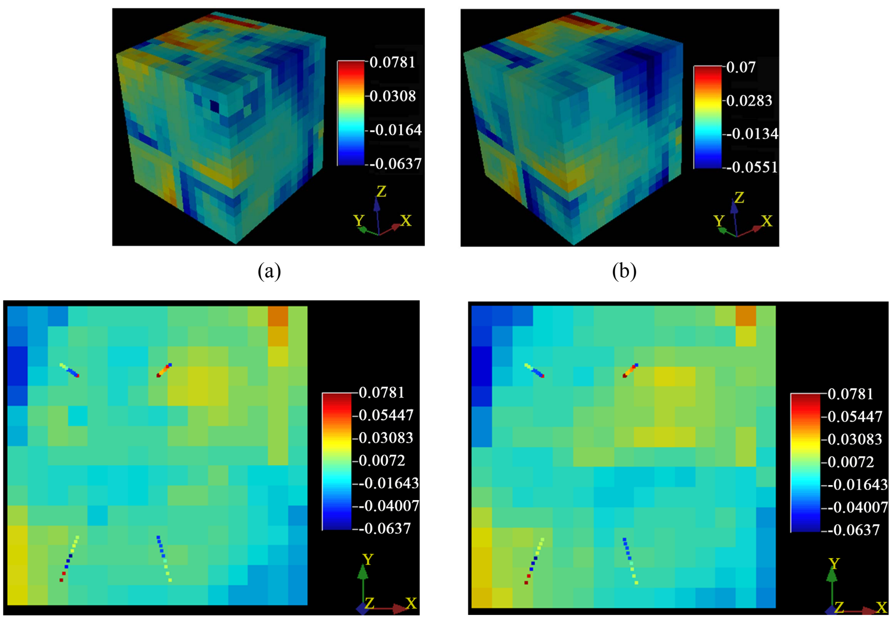


Figure 7. Simulation results. (a) Difference Diagram of Simulation Results between DCSK and SK; (b) Difference map of simulation results between FDSK and SK; (c) Differential Slice Map of DCSK in Z Direction; (d) Differential slice of FDSK in Z direction
图 7. 模拟结果差异图。(a) DCSK 与 SK 模拟结果差异图；(b) FDSK 与 SK 模拟结果差异图；(c) z 方向上 DCSK 的差异切片图；(d) z 方向上 FDSK 的差异切片图

7. 结论

当采用克里金对呈有限条带状分布的数据进行插值时，易产生条带效应，致使位于条件数据系列中两个端点的权值明显偏大。本文通过有限域克里金方法校正了克里金估值中的条带效应这种现象。有限域克里金方法是根据不同搜索邻域的最优克里金估量的平均值来得到估计值，是一种无偏的、高精度的估值方法，距离约束克里金方法与其相同，均是一种局部的校正方法。由于有限域克里金方法的每次估计都是最优的，不会产生不平滑的权值，因此相比于距离约束克里金方法，有限域克里金方法具有更好的校正效果。有限域克里金方法具有很好的稳定性，在条件数据个数不同的情况下都能表现出很好的权值修正效果。

基金项目

国家自然科学基金(41572121)、国家科技重大专项(2016ZX05033003)和湖北省自然科学基金创新群体项目(2016CFA024)联合资助。

参考文献

[1] Deutsch, C.V. and Journel, A.G. (1998) GSLIB: Geostatistical Software Library: And User's Guide. 2nd Edition, Oxford University Press, New York.

[2] Deutsch, C.V. (2002) Geostatistical Reservoir Modeling. Oxford University Press, New York.

[3] Deutsch, C.V. (1994) Kriging with Strings of Data. *Mathematical Geology*, **26**, 623-638. https://doi.org/10.1007/BF02089245

[4] Babak, O. (2008) Theoretical and Practical Improvements to Geostatistical Estimation and Simulation. University of Alberta, Edmonton.

[5] 郑光辉, 黄克龙, 张志宏, 等. 运用克里金空间插值技术进行土地级别划分[J]. 南京师大学报(自然科学版), 2007, 30(1): 112-116.

[6] 杜国明, 汪光松, 吴超羽, 等. 克里金在珠江河道地形空间数据内插中的应用[J]. 中山大学学报(自然科学版), 2007, 46(1): 119-122.

[7] 杨功流, 张桂敏, 李士心. 泛克里金插值法在地磁图中的应用[J]. 中国惯性技术学报, 2008, 16(2): 162-166.

[8] 陈亮, 黄述旺, 赖泽武. 应用截断高斯随机模拟法研究沉积微相[J]. 石油勘探与开发, 1998, 25(6): 78-80.

[9] 纪发华, 熊琦华. 序贯指示建模方法在枣南油田储层非均质研究中的应用[J]. 石油学报, 1994, 15(增刊): 179-186.

[10] 徐景帧, 刘晓冬, 等. 渗透率建模的三维指示条件模拟方法及其应用[J]. 石油学报, 2000, 21(3): 61-65.

[11] 李少华, 张昌民, 尹艳树. 储层建模算法剖析[M]. 北京: 石油工业出版社, 2012: 36-37.