Improved Factorial Kriging for Feature Identification and Extraction

Sahyun Hong and Clayton V. Deutsch

Centre for Computational Geostatistics
Department of Civil and Environmental Engineering
University of Alberta

Factorial kriging is a technique that aims to either extract features for separate analysis or mask out features from spatial data. Factorial kriging has been applied in several earth science application, however, most of the works were based on conventional factorial kriging which adopts ordinary kriging paradigm. This ordinary factorial kriging has strong constraint that forces sum of weights for the specific factor to be zero. In this work, we proposed an improved factorial kriging to extract spatial components. Factorial simulation was implemented to correct kriging smoothness and noise filtering of seismic data was tested using the proposed method. Factor data integration was initially suggested in this paper which integrates more relevant secondary factor with primary target variable. The considered method and applications were tested based on synthetic examples.

Introduction

The kriging interpolation algorithms aim to estimate the unsampled value of a variable in interest $Z(\mathbf{u})$. To apply kriging algorithm the spatial structures should be modeled, which is represented by variogram or equivalent covariance. When dealing with nested variogram or covariance structures it may be of interest to extract specific structures such as isotropic short range feature, anisotropic long range feature or nugget effect. Such filtering that excludes undesired feature and enhance the interesting feature can be achieved by factorial kriging. The number of sub-features and the magnitude of that features are chosen from the modeled nested variogram.

In this paper, we tested factorial kriging to identify the feature and to enhance target feature. Conventional factorial kriging is first considered. Conventional factorial kriging is based on ordinary kriging that assumes mean of $Z(\mathbf{u})$ is unknown over domain. Each factor is extracted using ordinary factorial kriging and the results are discussed. Some drawbacks of the method are observed even though ordinary factorial kriging is theoretically right. An improved factorial kriging method is proposed and tested using the same example. We observed the significant difference from conventional ordinary factorial kriging. Factorial kriging has smoothing effect so as simple kriging. We proposed factorial simulation that is conditional factorial simulation to correct smoothness of estimated factor.

Improved factorial kriging is also applied to filtering the noise inherent in exhaustive seismic data. Nugget effect is due to measurement errors hence it should be removed. Besides, nugget effect is also one structure consisting of the nested variograms. Thus, factorial kriging system of equations can filter out the unwanted nugget structure through excluding nugget effect model in the right hand side covariance term of kriging equation (see the details in following section).

Another application of factorial kriging is to integrate only relevant features extracted from original data. For some cases, global variability (large scale feature) is more important than local variability. In that case, large scaled feature is extracted from secondary data using factorial kriging and then primary variable and extracted relevant factor are co-kriged to estimate large scaled factor.

The applicability of the suggested method and application is tested using synthetic examples and effectiveness is discussed.

Setting

Paper 404 in this volume introduces conventional factorial kriging and modified `kt3d.exe` program. This paper follows the notation and explanation introduced in paper 404. Regionalized variable $Z(\mathbf{u})$ variable consists of a sum of independent factors and a mean:

$$Z(\mathbf{u}) = \sum_{l=0}^{L} Z_l(\mathbf{u}) + m(\mathbf{u}) = \sum_{l=0}^{L} a_l \bullet Y_l(\mathbf{u}) + m(\mathbf{u}) \tag{1}$$

The $L+1$ standard factors $Z_l(\mathbf{u}) = a_l Y_l(\mathbf{u})$ all have a mean of 0 and a variance of 1. The a_l parameters are stationary, which means they do not depend on location. The mean and variance of the $Z(\mathbf{u})$ variable are given by:

$$E\{Z(\mathbf{u})\} = E\left\{ \sum_{l=0}^{L} a_l \bullet Y_l(\mathbf{u}) + m(\mathbf{u}) \right\} = \sum_{l=0}^{L} a_l \bullet E\{Y_l(\mathbf{u})\} + m(\mathbf{u}) = m(\mathbf{u})$$

$$Var\{Z(\mathbf{u})\} = \sum_{l=0}^{L} a_l^2 \bullet Var\{Y_l(\mathbf{u})\} + 0 = \sum_{l=0}^{L} a_l^2$$

The variance of $Z(\mathbf{u})$ follows such a simple expression because $m(\mathbf{u})$ is a constant and the Y factors are standard and independent. These characteristics of m and Y also lead to a straightforward expression for the variogram of the $Z(\mathbf{u})$ variable:

$$\gamma\left(\mathbf{h}\right) = \sum_{l=0}^{L} a_l^2 \bullet \gamma_l\left(\mathbf{h}\right)$$

The $Z(\mathbf{u})$ regionalized variable is fully specified by $m(\mathbf{u})$, the $L+1$ a_l values, and the $L+1$ variograms $\gamma_l(\mathbf{h})$. The unknown parameter a_l^2 is magnitude of each nested structure. As discussed in paper 404 in this volume, we have access to original data values $Z(\mathbf{u})$ at sample location and the modeled variogram consisting of sub-structure. We do not know parameter a_l that specifies the importance of each factor. We do not have measurement of the factor $Y_l(\mathbf{u})$ either. We are able to distinguish the factors if the constituent variograms $\gamma_l(\mathbf{h})$ are different from one another. The reasonableness of factorial kriging depends entirely on the fitted nested structures. Therefore, when adopting a decomposition of sub-factors it is necessary to take account into physical interpretation and information.

Ordinary Factorial Kriging (OFK)

Conventional factorial kriging means an ordinary factorial kriging using the decomposition (1). The variable $Z(\mathbf{u})$ consists of factors $Z_l(\mathbf{u})$ and locally unknown mean $m(\mathbf{u})$. Consider the problem of estimating the spatial component $Z_l(\mathbf{u})$ of the decomposition. The OK estimator of that spatial component is

$$Z_l^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_{\alpha,l}^{OK} Z(\mathbf{u}_\alpha)$$

where $\lambda_{\alpha,l}^{OK}$ is the weights for the estimation of l component. It is noted that l factor $Z_l^*(\mathbf{u})$ is estimated using original data $Z(\mathbf{u}_\alpha)$, which includes $(L+1)$ factors and only the weights are different being associated with l factor. The estimation error of l factor is

$$E\{Z_l^*(\mathbf{u}) - Z_l(\mathbf{u})\} = \sum_{\alpha=1}^{n} \lambda_{\alpha,l}^{OK} m(\mathbf{u})$$

The constraint $\sum_{\alpha=1}^{n} \lambda_{\alpha,l}^{OK} = 0$ satisfies the unbiasedness condition. The error variance of l factor is expressed as double linear sum being similar to simple kriging

$$\sigma_{error,l}^{2}(\mathbf{u}) = Var\{Z_{l}^{*}(\mathbf{u})\} + Var\{Z_{l}(\mathbf{u})\} - 2Cov\{Z_{l}^{*}(\mathbf{u}), Z_{l}(\mathbf{u})\}$$

$$= \sum_{\alpha=1}^{n}\sum_{\beta=1}^{n}\lambda_{\alpha,l}\lambda_{\beta,l}C(\mathbf{u}_{\alpha}-\mathbf{u}_{\beta}) + C_{l}(0) - 2\sum_{\alpha=1}^{n}\lambda_{\alpha,l}Cov\{Z(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\}$$

The last term $Cov\{Z(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\}$ is reduced as

$$Cov\{Z(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\} = C_{l}(\mathbf{u} - \mathbf{u}_{\alpha})$$

since l factors, $l = 0,…,L$, are independent each other and since covariance can be shown as summing up of all factor covariance such as,

$$Cov\{Z(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\} = \sum_{l'=0}^{L}Cov\{Z_{l'}(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\}$$

$$\text{and } Cov\{Z_{l'}(\mathbf{u}_{\alpha}), Z_{l}(\mathbf{u})\} = \begin{cases} 0, & \text{if } l' \neq l \\ C_{l}(\mathbf{u} - \mathbf{u}_{\alpha}), & \text{otherwise} \end{cases}$$

The weights associated with l factor are obtained by minimizing the error variance under the unbiasedness constraint:

$$\begin{cases} \sum_{\beta=1}^{n}\lambda_{\beta,l}C(\mathbf{u}_{\alpha}-\mathbf{u}_{\beta}) - \mu_{l} = C_{l}(\mathbf{u}-\mathbf{u}_{\alpha}) \\ \sum_{\beta=1}^{n}\lambda_{\beta,l} = 0, & l = 0,...,L, \text{ and } \alpha = 1,...,n \end{cases}$$

(2)

The mean is estimated by linear combinations of the original data:

$$m^{*}(\mathbf{u}) = \sum_{\alpha=1}^{n}\lambda_{\alpha,m}\bullet Z\left(\mathbf{u}_{\alpha}\right)$$

The constraints on the weights to ensure unbiasedness:

$$\sum_{\alpha=1}^{n}\lambda_{\alpha,m} = 1$$

The estimation variances are minimized subject to the constraints given in the above leading to the factorial kriging equations to estimate the mean:

$$\sum_{\beta=1}^{n}\lambda_{\beta,m}C(\mathbf{u}_{\alpha}-\mathbf{u}_{\beta}) - \mu_{m} = 0, \ \alpha = 1,...,n \text{ and } \sum_{\beta=1}^{n}\lambda_{\beta,m} = 1$$

There are two interesting phenomenon to note. Ordinary factorial kriging system is very similar to ordinary kriging system except the right hand side covariance term. To estimate l factor the left hand side does not change but the right hand side should only consider the corresponding l^{th} covariance term. The resulting weights are interpreted as how much contribution of l factor in data value $Z(\mathbf{u}_{\alpha})$. The other phenomenon is that estimated component $Z_{l}^{*}(\mathbf{u})$ with constraint $\sum_{\beta=1}^{n}\lambda_{\beta,l} = 0$ has too small estimation variance which

results in smoothness, say, $\mathrm{var}\{Z_l^*(\mathbf{u})\} << \mathrm{var}\{Z_l(\mathbf{u})\}$. Traditional factorial kriging is assumed to have locally constant unknown mean, $m(\mathbf{u})$, however, there is no obvious reason why simple kriging could not be used.

Simple Factorial Kriging (SFK)

Many researchers have used factorial kriging to extract specific features or to filter out noise; however, most of the works were based on ordinary factorial kriging that addresses unknown mean. We proposed a factorial kriging with a constant known mean value. Locally constant unknown mean $m(\mathbf{u})$ is assumed to globally constant known mean $m(\mathbf{u}) = m$, which conforms to simple kriging paradigm. Simple factorial kriging equations with no constraints are derived with minimizing the error variance:

$$\begin{cases} \sum_{\beta=1}^{n} \lambda_{\beta,l} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_l(\mathbf{u} - \mathbf{u}_\alpha) \\ \alpha = 1,...,n \text{ and } l = 0,...L \end{cases} \tag{3}$$

Lagrange parameter that account for the non-bias constraint μ_l is removed in the simple factorial kriging.

The simple factorial kriging system of equation is similar to simple kriging equations except the right hand side. Interestingly it is noted that simple kriging weights $\lambda^{SK}(\mathbf{u})$ is equivalent to the sum of factorial simple kriging weights at the location. Let's see the systems of simple kriging equation,

$$\sum_{\beta=1}^{n} \lambda_\beta^{SK} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C(\mathbf{u} - \mathbf{u}_\alpha), \quad \alpha = 1,...,n$$

The RHS can be decomposed into $l = 0,...,L$, factors then,

$$C(\mathbf{u} - \mathbf{u}_\alpha) = \sum_{l=0}^{L} C_l(\mathbf{u} - \mathbf{u}_\alpha)$$

$$\sum_{\beta=1}^{n} \lambda_\beta^{SK} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = \sum_{l=0}^{L} C_l(\mathbf{u} - \mathbf{u}_\alpha) = \sum_{l=0}^{L} \sum_{\beta=1}^{n} \lambda_{\beta,l} C_l(\mathbf{u}_\alpha - \mathbf{u}_\beta)$$

since simple factorial kriging $\sum_{\beta=1}^{n} \lambda_{\beta,l} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_l(\mathbf{u} - \mathbf{u}_\alpha)$.

Thus, we have the following equality,

$$\sum_{l=0}^{L} \lambda_{\beta,l} = \lambda_\beta^{SK}, \qquad \beta = 1,...,n$$

,which represents simple kriging estimation is equal to the sum of simple factorial estimate, say

$$Z^{SK}(\mathbf{u}) = m + \sum_{l=0}^{L} Z_l^*(\mathbf{u}) \tag{4}$$

Factorial Simulation with SFK

Ordinary factorial kriging has inherent smoothing effect with sparse data and estimated mean factor explains the majority of the variance even with exhaustive measurements (discussed in Example). Simple factorial kriging relax the unbiasedness constraints, that is $\sum_{\alpha=1}^{n} \lambda_{\alpha,l} = 0$, $l = 0,...,L$, but simple factor estimate has smoothness effect, too. One possibility to correct the smoothness is to adopt simulation.

Every factor has different smoothness so that simulation performed factor-by-factor basis. Data values at sample location are used as conditioning information. Conditional simulation is worked based on the following scheme:

$$Z_{cs}(\mathbf{u}) = Z^*(\mathbf{u}) + [Z_{ucs}(\mathbf{u}) - Z^*_{uck}(\mathbf{u})]$$

where,

$Z_{cs}(\mathbf{u})$ is simulation value with conditioning original data $Z(\mathbf{u}_\alpha)$, $\alpha = 1,...,n$

$Z^*(\mathbf{u})$ is kriging estimate

$Z_{ucs}(\mathbf{u})$ is unconditional simulation value

$Z^*_{uck}(\mathbf{u})$ is kriging estimate using unconditional simulated values at data location

The reason factorial simulation can be generated using the above equation is that error term, $[Z_{ucs}(\mathbf{u}) - Z^*_{uck}(\mathbf{u})]$ can be generated for the specific factor l, $l = 0,...,L$ and kriging estimate term $Z^*(\mathbf{u})$ can be obtained from summing up of estimated factors, $\sum_{l=0}^{L} Z_l^*(\mathbf{u})$. Thus, conditional simulation for l^{th} factor can be obtained by

$$Z_{l,cs}(\mathbf{u}) = Z_l^*(\mathbf{u}) + [Z_{l,ucs}(\mathbf{u}) - Z^*_{l,uck}(\mathbf{u})] \tag{5}$$

In simulation, it should be checked that conditional simulation value at data location reproduces original data value at that location.

$$Z_{cs}(\mathbf{u}_\alpha) = Z(\mathbf{u}_\alpha), \; \alpha = 1,...,n$$

This checking method is slightly changed into such as sum of conditional factorial simulation values should reproduce data value at data location because we do not know true l factor, $Z_l(\mathbf{u})$ at that location.

$$\sum_{l=0}^{L} Z_{l,cs}(\mathbf{u}_\alpha) = Z(\mathbf{u}_\alpha), \; \alpha = 1,...,n$$

For arbitrary l factor, the error term $[Z_{l,ucs}(\mathbf{u}_\alpha) - Z^*_{l,uck}(\mathbf{u}_\alpha)]$ goes to 0 at data location due to the exactitude of kriging. Conditional simulation of l factor at data location $Z_{l,cs}(\mathbf{u}_\alpha)$ is $Z_{l,cs}(\mathbf{u}_\alpha) = Z_l^*(\mathbf{u}_\alpha) + 0$. Sum of conditional simulated factors at data location is equivalent to sum of simple kriged factor.

$$\sum_{l=0}^{L} Z_{l,cs}(\mathbf{u}_\alpha) = \sum_{l=0}^{L} Z_l^*(\mathbf{u}_\alpha)$$

We have shown that simple kriged value at location $\mathbf{u}$ is equal to sum of estimated factors at that location. Simple kriged value at data location is equal to data value itself due to exactitude of kriging hence,

$$\sum_{l=0}^{L} Z_l^*(\mathbf{u}_\alpha) = Z^{SK}(\mathbf{u}_\alpha) = Z(\mathbf{u}_\alpha), \; \mathbf{u}_\alpha \text{ indicates data location}, \; \alpha = 1,...,n$$

The above conditional factorial simulation scheme is checked.

Geostatistical Filtering Applied to Exhaustive Data

Random noise in exhaustive seismic data may be defined as noise which does not have a correlation between one point and its neighbouring points a short distance away. Random noise appears in exhaustive seismic data as an apparently random placement of bright or dark pixels and obscures fine details. There are several kinds of noise depending on source, but all can be placed in two categories; source-generated and ambient noise. Source-generated noise can take many forms, from surface waves to multiples to direct

arrivals and head waves. A typical cyclic noise appears in the 2-D or 3-D seismic data in form of repeated stripes due to designing of repeated seismic survey. Ambient noise can be caused artificially or naturally such as man-made noise, animal, air currents, nearby power lines. This undesirable noise is purely random and shown as salt-and-pepper in the seismic image. They must be removed to enhance original signal and identify the features.

As shown in kriging system of equations, factorial kriging filters out the designated structures. Long scale structures are removed when short range covariance is only remained in the right hand side of kriging equation and short scale structures are removed when long range covariance is only remained. Nugget effect is also treated within very short distance as one structure consisting of the whole nested covariance. Nugget effect is partly due to measurement errors; therefore, it might be better to filter out the nugget effect. Let us assume there are three factors noted as $l = 0, 1, 2$. 0th factor is usually considered as nugget effect that is to be removed. Filtered seismic map is achieved by the following factorial kriging system of equations

$$\sum_{\beta=1}^{n} \lambda_{\beta,l=1 \text{ and } 2} C(\mathbf{u}_\alpha - \mathbf{u}_\beta) = C_{l=1}(\mathbf{u} - \mathbf{u}_\alpha) + C_{l=2}(\mathbf{u} - \mathbf{u}_\alpha), \alpha = 1,...,n$$

$$Z_{Filtered}(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_{\alpha,l=1 \text{ and } 2} Z(\mathbf{u}_\alpha)$$

Geostatistical filtering applied to seismic data is represented in the following example section.

Data Integration

One potential application of factorial kriging is data integration with more relevant features. Key concept is to filter out undesirable noise and artifact in the secondary data and extract relevant feature or factors from cleaned secondary data. Factorial kriging of primary variable is then performed using relevant factor extracted from secondary data. For example, large scale feature is important rather than locally variable feature in a certain case. In that case, large scale feature is extracted from the secondary data(already filtered the nugget effect) which filters out short range feature using simple factorial kriging. The extracted secondary feature is then used as a secondary data in order to estimate primary variable, that is:

$$Z_l^*(\mathbf{u}) = \sum_{\alpha=1}^{n} \lambda_{\alpha,l} Z(\mathbf{u}_\alpha) + \mu_l Y_l(\mathbf{u}) \qquad (6)$$

where $Z_l^*(\mathbf{u})$ is estimated primary variable for l factor.

$\lambda_{\alpha,l}$ are the weights assigned for data samples to estimate l factor.

$Z(\mathbf{u}_\alpha)$ is data samples.

$Y_l(\mathbf{u})$ is the extracted relevant secondary factor at estimation location $\mathbf{u}$.

μ_l is the weight assigned for secondary factor $Y_l(\mathbf{u})$ at estimation location $\mathbf{u}$.

This method may be called factorial collocated co-kriging since only collocated secondary factor is retained to estimate. The weights are obtained by solving collocated co-kriging equations for l factor.

$$\sum_{\beta=1}^{n_1} \lambda_{\beta,l}(\mathbf{u}) C_{ZZ}(\mathbf{u}_\alpha - \mathbf{u}_\beta) + \mu_l C_{ZY_l}(\mathbf{u}_\alpha - \mathbf{u}) = C_{ZZ}^l(\mathbf{u}_\alpha - \mathbf{u}), \ \alpha = 1,...,n$$

$$\sum_{\beta=1}^{n_1} \lambda_{\beta,l}(\mathbf{u}) C_{Y_l Z}(\mathbf{u} - \mathbf{u}_\beta) + \mu_l C_{Y_l Y_l}(0) = C_{Y_l Z}(0)$$

$$(7)$$

Example

The first synthetic example is shown to show how factorial kriging approach works. A synthetic 256×256 2-D Gaussian variable was simulated with a small nugget effect (10%) and two isotropic spherical structures equally explaining the remaining 90% of the variability. The ranges of the spherical structures are 16 and 64 units. Data were sampled from the reference grid at a very close 5×5 spacing to produce sample data. The reference grid and sample data are shown on Figure 1.

As developed above, conventional factorial kriging is based on ordinary kriging. The sum of the estimate of each factor adds up to the ordinary kriging estimate. The ordinary kriging estimates are shown on Figure 2. Note that the ordinary kriging estimates are smoother than the reference values – a characteristic property of all kriging. Note also that the data are reproduced exactly, but with an apparent discontinuity because of the nugget effect.

The ordinary factorial kriging estimates the mean and each factor independently. In this example there are the mean and three factors to be estimated. Maps of the factors are shown on Figure 3. Although the nugget effect map looks constant, it is not. There is a discontinuity at each data point. Note that the estimate of the mean reflects the most variability because of the unbiasedness constraints which forces sum of weights to be 0 for the considered factor. Although the kriging equation and unbiasedness condition is right, it is unrealistic that estimated mean factor has the majority of variability. Furthermore, estimated long range structure is destructed as shown in Figure 3 and has low variability.

Simple factorial kriging estimates of each factor are shown in Figure 4. The mean value is assumed to be constant overall. Short range factor estimates have variability within short distance and long range factor estimates have variability within large distance in terms of visual interpretation. Simple factorial estimates reflect realistic features contrary to the ordinary factorial kriging estimates. Short range feature and long range features are identified significantly. Over every estimation location, sum of all estimated factors reproduced simple kriged value as we have theoretically derived equation (4).

OFK vs SFK

It is expected to show each estimated factor has equal or close to the variability observed in the nested variogram. Table 1 represents variance of each estimated factor by OFK and SFK, respectively. This variance table represents how much different between OFK and SFK estimates quantitatively. The numeric values show the calculated variance for each factor. First column is the contribution of each factor obtained by nested variogram modeling. Second and third column show the variance of each estimated factor with OFK and SFK. The last column in Table 1 shows how much OFK smooth the estimated factor being compared to SFK. The estimated mean factor using OFK has 70% variability which is the majority of variability. From the table, SFK has more variability than OFK estimate and SFK produces more reasonable results in terms of feature identification. However, smoothing effect still exits.

Table 1. Variance table for each estimated factor by OFK and SFK

Factors	$b_l = (a_l^2)$	OFK	SFK	diff in %
mean	–	0.7	0	70 %
$l = 0$ **(nugget)**	0.1	0.0	0.0	0 %
$l = 1$ **(short range)**	0.45	0.15	0.2	25 %
$l = 2$ **(long range)**	0.45	0.01	0.42	95 %
Sum of all factors	1.0	0.97	0.99	2%

Factorial Simulation

Factorial simulation with SFK is worked using the same synthetic data. Three factors are considered; nugget factor, isotropic small scale and isotropic large scale factors. We followed the conditional factorial

simulation in equation (5) to adopt conditioning information. First of all, unconditional simulations were generated for each factor. Left column in Figure 5 represents one unconditional realization for three factors. l^{th} factor covariance, $C_l(\mathbf{h})$ was only used when generating l^{th} unconditional simulation instead of using whole covariance $C(\mathbf{h})$. The right column in Figure 5 shows the resulting conditional simulation. 0th factor(nugget effect) simulation seems constant overall, but they have values with very low variance. Table 2 is same as Table 1 except inserting the last column for conditional simulation results. For the first realization, we calculated variance of each estimated factor and observed smoothness effect inherent in all kriging estimate is reduced.

Table 2. Variance table for each estimated factor with OFK, SFK, and conditional simulation using SFK. The first realization is selected here.

Factors	$b_l = (a_l^2)$	OFK	SFK	Cond'l Simulation
mean	–	0.7	0	0
$l = 0$ **(nugget)**	0.1	0.0	0.0	0.1
$l = 1$ **(short range)**	0.45	0.15	0.2	0.39
$l = 2$ **(long range)**	0.45	0.01	0.42	0.44
Sum of all factors	1.0	0.97	0.99	1.0

The unconditional and conditional simulations have interesting relationships with respect to different factors. The unconditional simulation can be viewed as being to be updated through adopting conditioning data. The contribution of conditioning information is more significant when simulating large scale factor(Factor 3 in this example) as shown in the third row in Figure 5. Unconditional simulation of small scale factor(Factor 2 in this example) is not much influenced by conditioning data as shown in the second row in Figure 5. Figure 6 represents the amounts of change in simulation values when adopting conditioning data and not. X-X' section was chosen for showing. Bottom graph in Figure 6 illustrates conditioning data(✚), unconditional simulation(solid line), and conditional simulation(dashed line) along with X-X' line. There is rare correlation between unconditional simulation and conditioning data points. After updating by the conditioning data, conditional simulation is likely to follow conditioning data points. We have observed large amount of updating after incorporating conditioning data for large scale feature. This phenomenon is not observed in small scale factor simulation.

It is noted that conditional simulation(dashed line) does not reproduce the conditioning data at data location instead sum of conditional simulated factors reproduce conditioning data.

Geostatistical Filtering

Noisy synthetic seismic data is prepared. Seismic amplitude map is generated over the channel type reservoir and amplitude signal is exaggerated to clearly show the effect of noise filtering. One can see meandering channel in the seismic amplitude image. Original seismic image with noise is shown in Figure 7. The modeled variogram is also shown in the first row in Figure 7. Solid lines represent fitted variogram with experimental variogram shown as dashed line. As expected, the data has high nugget effect($\approx 50\%$). The seismic data has three nested structures, i.e. 50% nugget effect, 30% isotropic small structure and 20% anisotropic structure with N-S major continuation direction. This example has no systematic noise such as repeated noise bands due to cyclic seismic survey design. Salt-and-pepper type random noise is only shown. Seismic amplitude is transformed into normal scores first. Our goal is to mask the unwanted random noise and to enhance the meaningful structures. Simple factorial kriging was applied to this

example. Second and third row in Figure 7 show extracted factors corresponding to nugget, isotropic short range, and anisotropic long range. Nugget effect is due to random noise so that we filtered out nugget factor and resulting filtered image is shown in the lower right in the same figure. Random noise is largely removed with preserving small feature and large feature.

Geostatistical filtering is basically depending on the number and types of fitted variogram. This is one advantage of geostatistical filtering technique since geologic knowledge can be added when fitting the experimental variogram hence more realistic structure may be extracted.

Data Integration

Amoco.dat data is considered for data integration. We have one primary porosity variable to be estimated and one exhaustively sampled seismic data. Location of primary porosity is shown in upper left of Figure 8. Normal score seismic amplitude is shown in right of the figure. Fitted variogram of porosity has three identified features, i.e. 10% nugget effect, 20% isotropic small scale and 70% anisotropic large scale features (major in N-S and minor in E-W).

$$\gamma(\mathbf{h}) = 0.1 + 0.2 Sph(\mathbf{h})_{\text{hmin=hmax=1000}} + 0.7 Sph(\mathbf{h})_{\substack{\text{hmin=6000} \\ \text{hmax=25000}}}$$

Seismic data appears to be varied smoothly. Variogram of secondary data is fitted as having two features, 35% isotropic small scale and 65% isotropic large scale features.

$$\gamma(\mathbf{h}) = 0.35 Sph(\mathbf{h})_{\text{hmin=hmax=2000}} + 0.65 Sph(\mathbf{h})_{\text{hmin=hmax=20000}}$$

First, simple factorial kriging with exhaustive seismic data was performed and two factors are extracted separately: isotropic small scaled feature and isotropic large scaled feature. Extracted factors are shown in Figure 9. To check the relationships between primary porosity and extracted factor, scatter plots are shown in the right column in Figure 9. Correlation of primary, $Z(\mathbf{u}_\alpha)$ and total seismic variable, $Y(\mathbf{u}_\alpha)$ is summarized as ρ_{ZY} and ρ_{ZY} is 0.615. It is noted that correlation coefficient is slightly changed when plotting extracted factor 1 and 2 with primary variable,

$$\rho_{ZY_{l=1}} = 0.488 < \rho_{ZY_{total}} = 0.615$$

$$\rho_{ZY_{l=2}} = 0.660 > \rho_{ZY_{total}} = 0.615$$

where $Y_{l=1}$ is an extracted small scale feature and $Y_{l=2}$ is an extracted large scale feature from original seismic data noted as Y_{total}. We checked that large scale factor is more relevant feature to the primary variable. This is reasonable result since the variogram of primary porosity has greater contribution of large scale feature (70% variability) than small scale feature (20% variability). Besides, the variogram of seismic data shows 65% variability contribution of large scale feature. Although the difference between $\rho_{ZY_{l=1}}$ and

$\rho_{ZY_{total}}$ or the difference between $\rho_{ZY_{l=2}}$ and $\rho_{ZY_{total}}$ is not large, it is important to check that extracted relevant factor is more related to the primary variable to be estimated. Figure 11 represents the integrated isotropic small scale feature and anisotropic large scale feature using factorial collocated co-kriging equation (7). To identify isotropic small scale feature, primary variable and extracted isotropic small scale factor are co-kriged. To identify anisotropic large scale factor, primary variable and extracted isotropic large scale factor are co-kriged. All co-kriged estimates are shown in Figure 11.

To see how much update after integration, small scale and large scale features extracted only from primary porosity are plotted in Figure 10. Anisotropic long range structure is significant along with N-S direction. The variance of small scale factor is very low due to smoothness effect. Co-kriged factor 1 and 2 has more variability as shown in Figure 11.

Conclusions

An improved factorial kriging algorithm has been proposed to identify features. Conventional factorial kriging method has been reviewed and tested on synthetic example, and they showed unrealistic estimation even though kriging equation and constraint are theoretically right. An improved factorial kriging adopts simple kriging paradigm. Unknown mean is assumed to be constant overall and simple factorial kriging was applied on the example data. Features with different scale are identified well rather than using conventional factorial kriging.

Factorial simulation with simple factorial kriging has been suggested to correct smoothness of kriging. Conditional factorial simulation worked based on combination of factorial kriging and unconditional factorial simulation. Unconditional factorial simulation was updated significantly in large scale feature after assigning conditioning data. Factorial simulation corrects the kriging smoothness.

Other application of factorial kriging was to mask noise in exhaustive seismic data and to amplify the specific spatial features. Synthetic seismic amplitude data was prepared with large random noise. Simple factorial kriging successfully filtered out the unwanted random noise and enhanced the specific spatial factors. Factorial kriging approach makes it possible to integrate more relevant feature with primary variable. We have tested factorial collocated co-kriging using sparely sampled porosity and extracted relevant factor from seismic data. After integrating, we have identified spatial features with different scales which are hidden in the primary variable.

References

Chiles J-P. and Delfiner P., 1999, *Geostatistics Modeling Spatial Uncertainty*, John Wiley & Sons, Inc.

Deutsch, C.V. and Journel, A.G., 1998, *GSLIB: Geostatistical Software Library: and User's Guide*. Oxford University Press, New York, 2nd Ed.

Goovaerts, P., 1997, Geostatistics for Natural Resources Evaluation, Oxford University Press, New York.

Deutsch, C.V., 2007 A Recall of Factorial Kriging with Examples and a Modified Version of kt3d, *Centre for Computatinal Geostatistics Annual Report 9*.
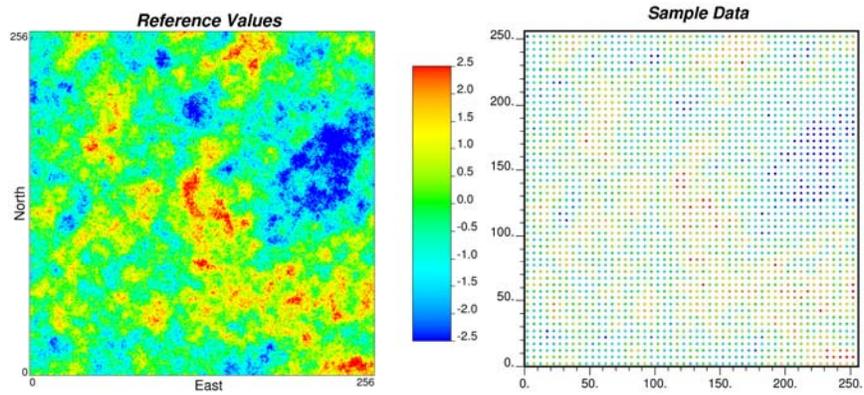
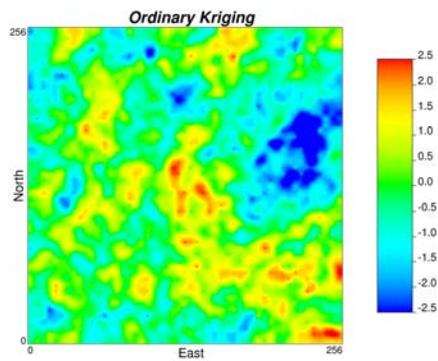**Figure 1**: Reference values and sample data for example.

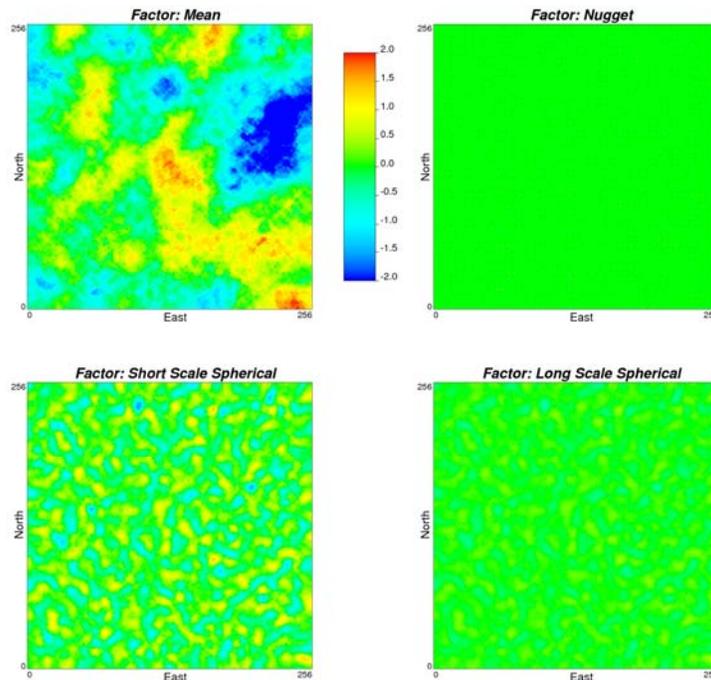**Figure 2**: Ordinary kriging for example.

Figure 3: Ordinary factorial estimates: mean, nugget, short scale and long scale features.
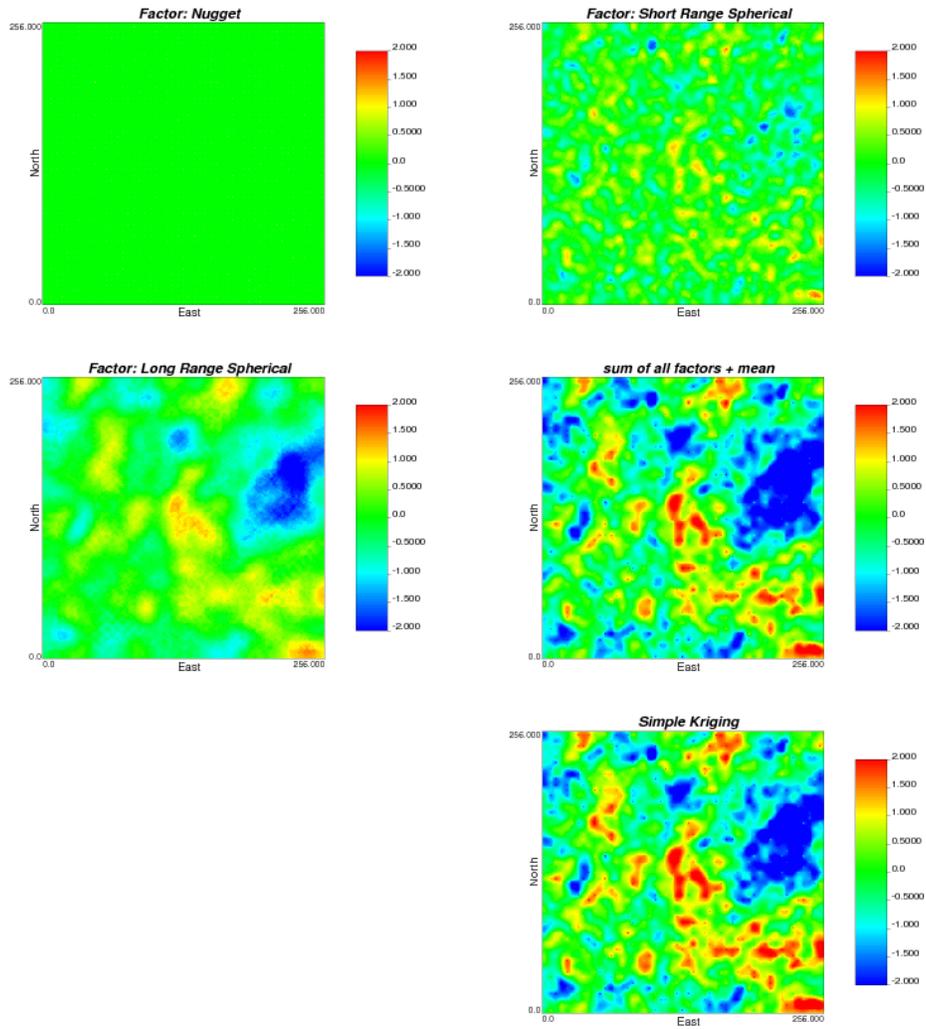
Figure 4: Simple factorial estimate. Sum of estimated factors and constant mean is very close to simple kriging estimate(shown in the bottom).
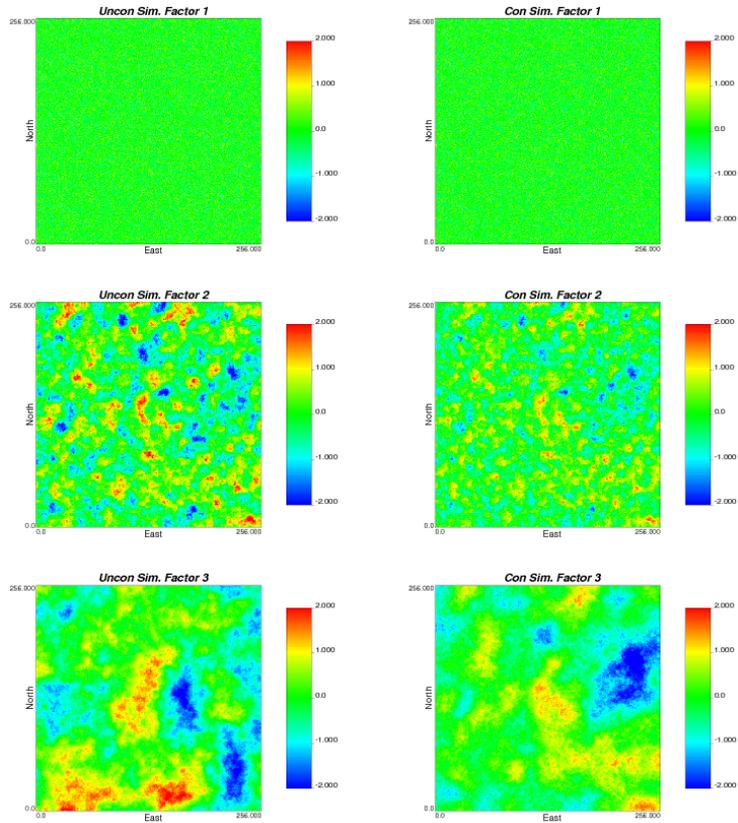
Figure 5: Unconditional and conditional factorial simulation using simple factorial kriging. Factor 1, 2, and 3 indicate nugget, short range and long range structures, respectively.
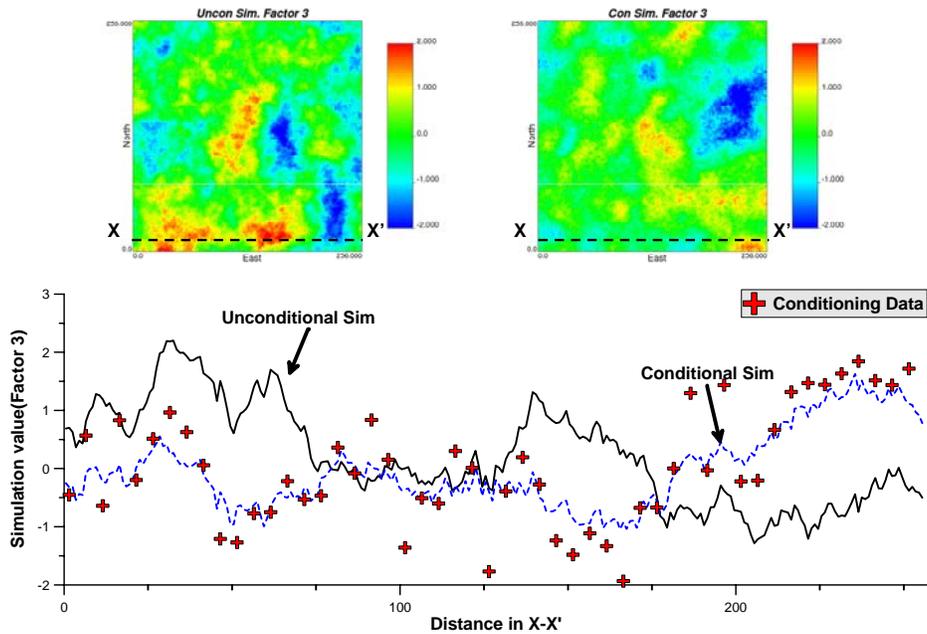
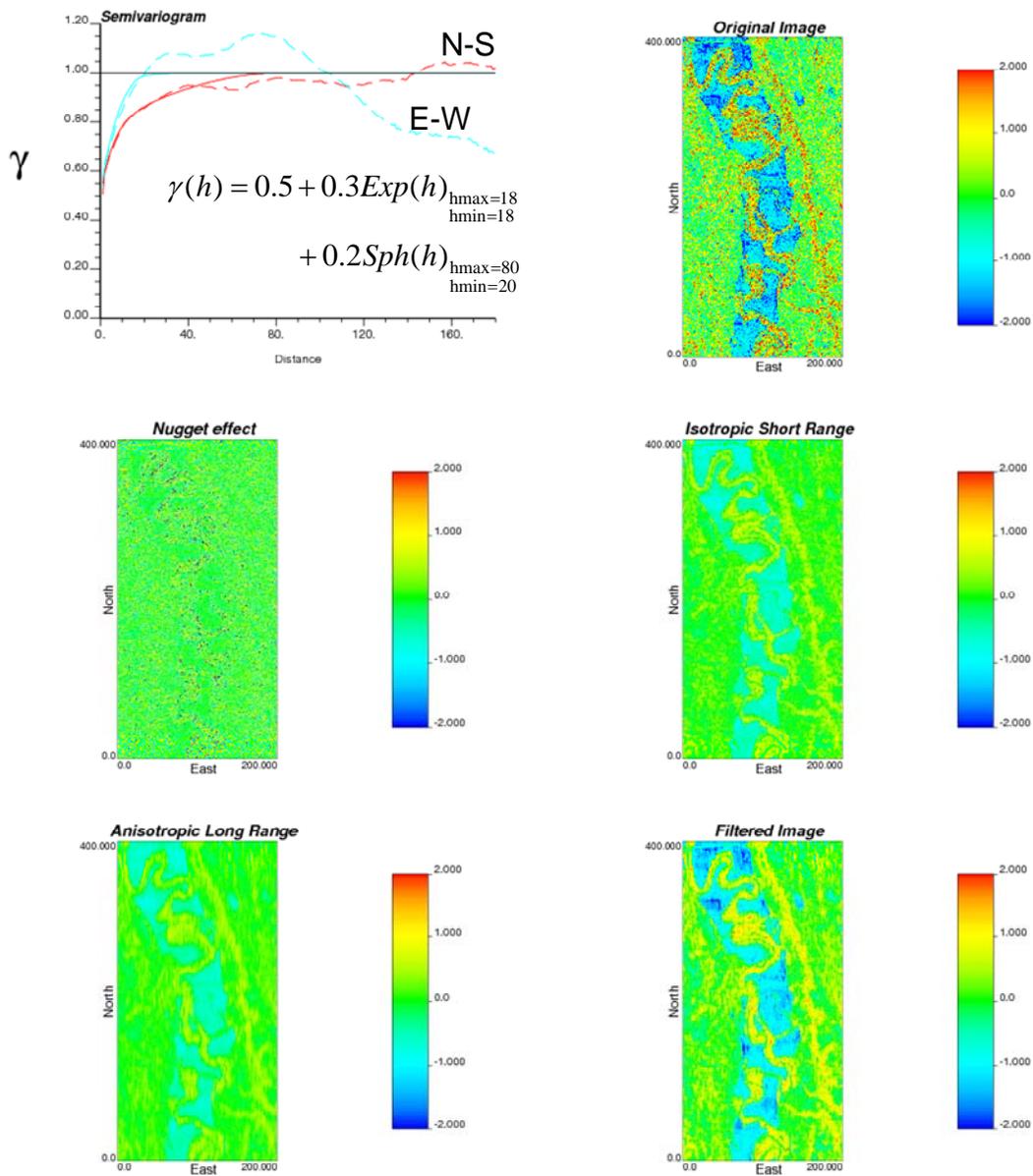**Figure 6:** Unconditional and conditional simulation values in line X-X'. The profile represents how much conditioning data affect the simulation with long range feature.

$$\gamma(h) = 0.5 + 0.3 Exp(h)_{\substack{hmax=18 \\ hmin=18}}$$
$$+ 0.2 Sph(h)_{\substack{hmax=80 \\ hmin=20}}$$

Figure 7: Geostatistical filtering applied to seismic exhaustive data. The original seismic amplitude is generated. Synthetic example contains large random noise that should be removed. Simple factorial kriging is applied to filter out the noise and to extract specific features.
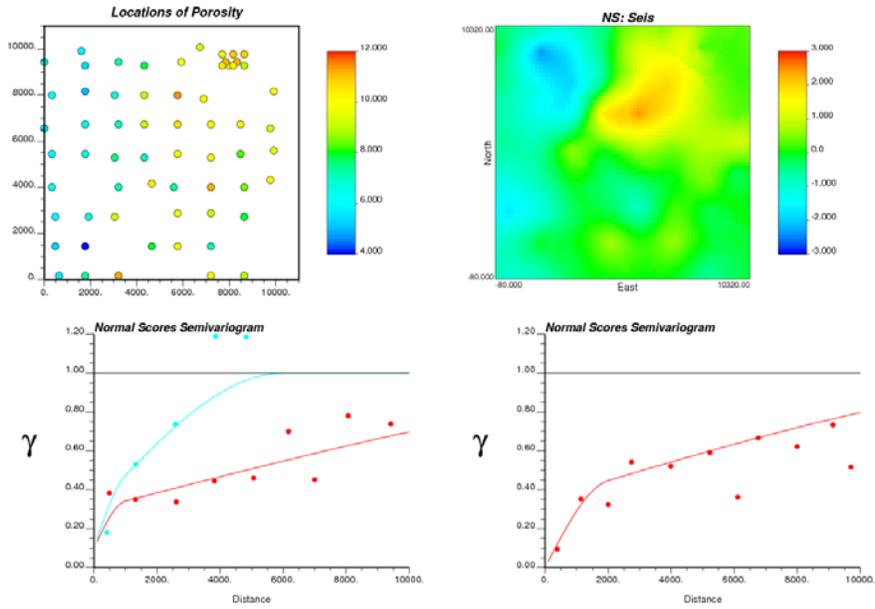
Figure 8: Primary and secondary variable map and their fitted variograms.
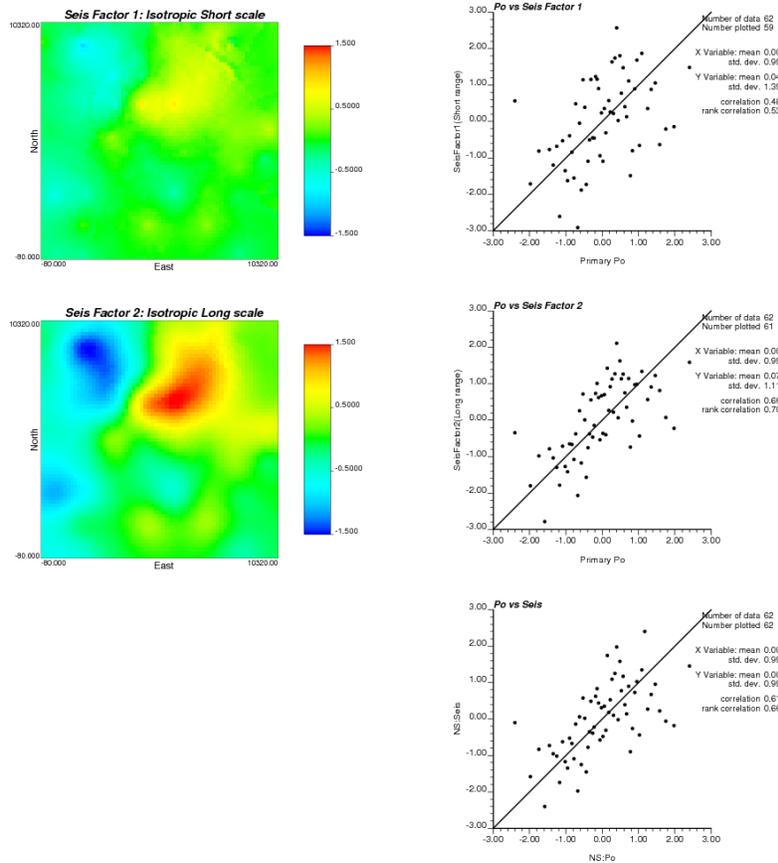

Figure 9: Extracted factor from secondary data. Right scatter plot shows the correlation between primary and extracted factor 1 and 2. The last scatter plot is the plot between primary and secondary.
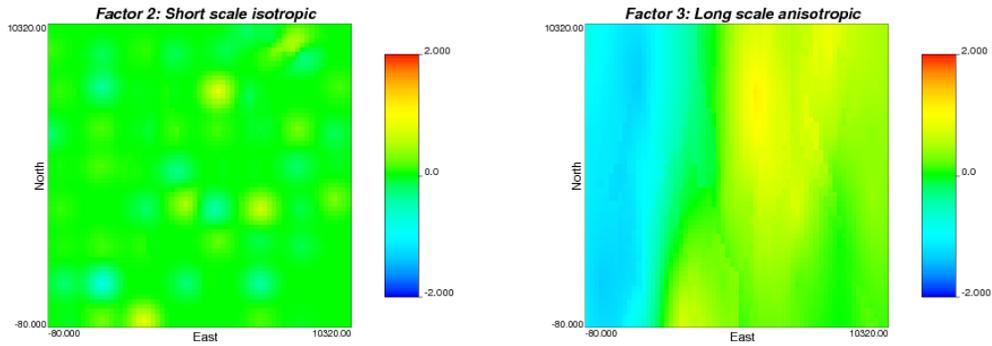
Figure 10: Extracted factor 1 and 2 from the primary variable. Primary variable has two structures, isotropic small feature and anisotropic (major continuity in N-S) long range feature.
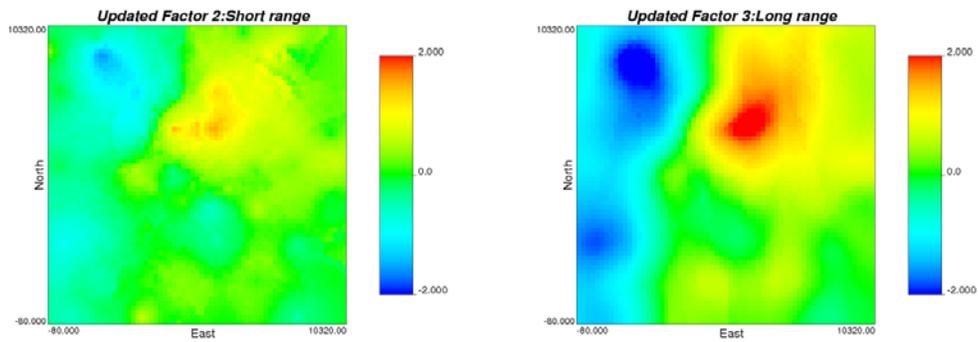

Figure 11: Integrated maps. Collocated co-kriging is performed with primary and relevant factors extracted from secondary data. Left figure is the integration map with primary and short range factor 1. Right figure is the integration map with primary and long range anisotropic factor 2.