

# Linear and Non Linear Estimation Based on Training Images

David F. Machuca Mory and Clayton V. Deutsch

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

*The weights in linear estimation can be optimized with kriging; however, a direct optimization procedure is considered here. The weights are optimized directly using a training image. The weights can depend on the data values or, as in kriging, be data value independent. Strings of data are considered to assess whether the weights to end samples should be larger – they should. Non-linear estimation can easily be considered in the direct calculation of weights; kriging would require high order covariances. All these results are compared with the sample weights obtained from Kriging. The training image derived weights are more flexible reflecting the spatial features of training images, especially when strong anisotropies and discontinuities are present, they lead to an improved estimation in terms of the optimization criteria chosen. The cost of this improved estimation is (1) the availability of a relevant training image, and (2) the CPU time to perform the optimization.*

## Introduction

Kriging based estimation techniques use the spatial correlation information provided by the variogram to calculate the weights of the sample values surrounding a point or block to be estimated. The sample weights obtained from the kriging equations minimize the estimation variance and accounts for the spatial correlation between the surrounding samples and the estimated point / block and between samples themselves. The variogram, however, fails to fully represent the complex spatial relationships.

An alternative is to directly employ optimization techniques to obtaining optimal weights from a pertinent and representative training image of the spatial distribution of that particular variable under study. Then, these weights can be used to estimate with the available data.

A simulated annealing-like steepest descent method is used in the present work. An original arrangement of weights is randomly perturbed in a recursive way, the changes in the weights are kept only if they lead to a lower estimation error under a well defined error criteria (squared error, absolute error or other criteria). This process requires a large number of iterations and training image for reference data and true values.

The optimized weights and errors using this approach are compared with the ordinary and simple kriging weights and error for different “true” training images. The characteristic features of the optimized weights are studied relative to the features of the training images values, such as Non-stationarity, Non-Gaussianity and discontinuities. The effect of non-stationarity when estimating in a finite domain is also studied using synthetic training images. Both data independent and dependent weights are considered; in the last case a non-linear estimator is considered.

## Methodology

The methodology for weights optimization followed in this work is similar to the one proposed by Deutsch, Ortiz and Bhandari (2006), this is an iterative procedure that takes the numerical values of a training image pixels in order to assemble a big number,  $M$ , of data and true values for a predetermined data arrangement. The original data weights are randomly perturbed and if no constraint is imposed to the sum of the weights the estimated value is calculated by a linear estimator of the form:

$$z^*(\mathbf{u}) - m(\mathbf{u}) = \sum_{i=1}^n \lambda_i \cdot [z(\mathbf{u}_i) - m(\mathbf{u}_i)]$$

Where  $u$  is a location within the training image,  $m$  is the mean of the pixel values in the training image,  $z^*(u)$  is the estimated value at location  $u$ ,  $z(u_i)$ . Otherwise, if this sum is constrained to 1 the linear estimator becomes:

$$z^*(u) = \sum_{i=1}^n \lambda_i \cdot z(u_i)$$

The weights of these linear estimators are independent of the data values. Data dependent weights can be defined as:

$$\lambda_i = \sum_{j=0}^m a_{ij} (z(u_i))^j$$

Where  $a_{ij}$  are the coefficients of a polynomial of order  $m$ . Therefore the non-linear estimator with data dependent weights becomes:

$$z^*(u) - m(u) = \sum_{i=1}^n \left[ \sum_{j=0}^m a_{ij} (z(u_i))^j \right] \cdot [z(u_i) - m(u_i)]$$

In this case the coefficients  $a_{ij}$  of each data point are slightly perturbed iteratively. In both cases, altering the weights or altering the polynomial coefficients, the changes are accepted if these lead to a lower value of the optimality criteria used, which can be, among the most important:

Mean Square Error (MSE) 
$$MSE = \frac{1}{M} \sum_{l=1}^M [z_l^* - z_l]^2$$

Least Absolute Deviations (LAD) 
$$LAD = \frac{1}{M} \sum_{l=1}^M |z_l^* - z_l|$$

The MSE is the most used optimality criteria in this work,  $M$  is the number of arrangements conformed by the samples locations and estimated points that can be defined within the training image.

### Estimating with strings of data

In Kriging, the string effect appears in the form of higher weights assigned to the samples located in the extremes of a string of data. This effect is observed in both Simple (SK) and Ordinary Kriging (OK), but is more pronounced in OK. It is controlled by the variogram ranges,  $a_x$ ,  $a_y$ , the distance from the estimated point to the string,  $d$ , the string length,  $L$ , the constant spacing between samples,  $s$ , and the number of samples in the string,  $n$ . (See Figure 1)

The central samples get higher weights when the ratio  $d/a_x$  decreases and the ratios  $s/a_y$  and  $L/a_y$  increases; this behaviour is induced by the higher importance acquired by the correlation between the samples and the estimated point in the kriging matrix when the anisotropic distance to the estimated point decreases, the anisotropic separation between samples increase, as well as the anisotropic length of the string. The sample weights in the extremes follow nearly the inverse behaviour; being the ratios  $s/a_y$  and  $L/a_y$  the most important factor than control them. This can be explained by realizing that if the variogram range parallel to the string increases, the correlation between the samples in the string becomes more important than the correlation between the samples and the estimated point, thus the high weights assigned to samples in the

extremes reflect the importance that acquire the information in the direction parallel to the string, which is, however, limited to few samples.

Increasing the number of samples in the string without increasing its length leads to a redistribution of weights maintaining the same previous weights profile, the central samples can get slightly lower weights due to the lowering of the ratio  $s/a_s$ . If the number of samples is increased by adding new samples at the extremes of the string, and thus increasing the string length, the weights of the former samples in the ends are redistributed with the new ones and the central samples get higher weights because enlarging the string has the same effect as decreasing the distance to the estimated point.

*Obtaining optimal weights from different TI's for a string of data*

The string effect is also present when the sample weights are obtained from a TI by optimization. But in this case, the weights distribution profiles are not induced by the variogram model used but by the spatial relationships inherent to the TI. Moreover these weights profiles are not necessarily symmetrical, neither the string effect appear in them at the same anisotropic distances as it appears in kriging. Figures 3 and 4 show a comparison of the weights profiles at different distances for two training images using SK, OK, unconstrained and constrained optimization. The training images used for deriving the estimation weights are coded as p162f and p216z (the plots and histograms of these two training images used are shown in the appendix of this paper). The p162f TI is approximately stationary and show regular spatial features with constant angles and ranges of the anisotropy ellipsoid, the variogram model fitted to the experimental variogram obtained from the pixel values of this TI has the following parameters:

**Table 1:** TI p162f, fitted variogram model parameters

Model	Sill	Az	a_max	a_min
<b>Nugget Effect</b>	0.001			
<b>Gaussian</b>	0.312	90	24.5	7.2
<b>Gaussian</b>	0.554	90	65	9
<b>Gaussian</b>	0.133	90	455	14

By the contrary, the second training image, coded p216z, shows discontinuities with zones of different anisotropy orientation. The best variogram model fitted for this TI is defined by the next parameters:

**Table 2:** TI p216z, fitted variogram model parameters

Model	Sill	Az	a_max	a_min
<b>Nugget Effect</b>	0			
<b>Gaussian</b>	0.56	100	21.0	4.0
<b>Gaussian</b>	0.44	90	140.0	5.8

For a regular stationary TI image such as p162f, the unconstrained optimized weights profiles are fairly similar to those obtained with simple and ordinary kriging only at short distances. These profiles diverge quickly as the distance to the estimated point increases; the profiles obtained by optimization are more flexible reflecting the spatial features of the TI at different distances, as it can be appreciated in the systematic higher weights that are assigned to the top samples in the p162f TI (see Figure 3), or as it is shown in the weights obtained from the TI p216z (see Figure 4), where the short scale variogram produces slightly asymmetrical profiles for SK and OK at short distances while the optimal weights profiles show a stronger asymmetry at short distance but a more symmetrical shape at long distances, as well as important weights for the central sample at all distances. In this case, the weights profile asymmetry at short and mid distances and the symmetry at long distances can be attributed to the zones of different layering orientation and the discontinuities present in this particular training image.

This example highlights the “rigidity” of variogram based estimation and the variogram limitations for depicting the spatial features of the training images with enough detail. Figures 5a and 5b show the SK errors and the unconstrained optimization errors, and the OK errors and constrained optimization errors, respectively, for the p162f TI. Equivalent graphs are presented in Figures 5c and 5d for the p216z TI. There it can be observed that Kriging and optimization errors eventually reach a sill at longer distances for optimization than kriging minimum squared errors. This distance does not necessarily correspond to the variogram range but it is related to the capability of the variogram induced kriging weights to depict the

spatial features of the Training image at long distances. Similarly, the TI derived MSE's can also reach a sill at certain distance. The average MSE differential ( $\Delta E$ ) between the errors obtained using kriging and TI derived weights grows with the distance until reaching a peak and decreases at longer distances. The average weights difference between kriging and TI derived weights follows a similar path of growing until reaching a peak and becoming stable or decreasing at farther distances.

In general, the weights average difference and the average MSE differential have a positive relationship at a fixed pixel distance when several different training images are considered (see Figures 6a and 6b), this means that, usually, as the TI derived weights diverge from the kriging weights, the average MSE improvement is higher. However the correlation between these two measures is not very strong, this can be explained in regard to their behaviour description above and the particular spatial features of each training image.

### Non stationarity in finite domains

In order to study the response of TI weights in comparison to Kriging weights within a non stationary finite domain a set of synthetic training images was generated. First, a standard Gaussian residual was simulated by Sequential Gaussian simulation over a 50,000 x 9 pixel grid. The variogram used was a spherical model with 40 pixels range in the direction parallel to the X axis, and a range of 7 pixels in the Y axis direction, and the nugget effect was set to zero. Then, the composite variable in the final training images were generated by different combinations of vertical trends ( $T(u)$ ) multiplied by an slope factor ( $a_T$ ) plus residuals ( $Z(u)$ ) also multiplied by an amplitude factor ( $a_Z$ ). Figure 7 shows the four vertical trends used: linear, upper side linear, quadratic and cubic. Thus, the composite training image variogram in the vertical direction become:

$$\begin{aligned} 2\gamma(h) &= E \left\{ \left[ (a_T \cdot T(u) + a_Z \cdot Z(u)) - (a_T \cdot T(u+h) + a_Z \cdot Z(u+h)) \right]^2 \right\} \\ &= a_T^2 \gamma_T(h) + a_Z^2 \gamma_Z(h) \end{aligned}$$

Where  $\gamma_T(h)$  and  $\gamma_Z(h)$  are the variograms of the vertical trend and the residuals respectively. However, this formula leads to non permitted models if non linear trends are used. So the experimental variograms of the composite variable were calculated in the two major directions and approximately fitted using permitted models.

When estimating a point located a the middle of the band with a linear trend, the TI derived weights become very close to the Kriging derived weights using the variogram of residuals for both, constrained and unconstrained optimization (See Figures 8a and 8b). This can be explained by the fact that being the slope of the trend constant, the estimation weights at the middle of the band are mostly controlled by the horizontal component of the composite variogram above, which is just the variogram of the residuals. Consequently, the MSE obtained by both methods are very close (see table 3a).

This is not the case when the estimated point is located at the top of the band, then the correlation between the samples in the vertical string becomes as important as the correlation between samples and the estimated point, thus the whole composite variogram is needed. Figures 8c and 8d present the TI derived weight profiles in unconstrained and constrained optimization compared with SK and OK weights, respectively. In these Figures it can be observed that, for TI derived weights, increasing the residuals amplitude has a result a decreasing in the differentiation between border and middle samples weights and an approaching between TI and Kriging derived weights, which is concordant with the MSE presented in the table 3b.

It is also worth to note that, while SK and OK derived weights for the sample at the bottom of the string are very close to zero or show a slightly positive string effect, the TI derived bottom sample weight can show an inverse string effect, particularly in no constrained optimization.

**Table 3a:** MSE comparison for estimation at the middle of a band with linear vertical trend

Combination	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
2T+0.25Z	0.053	0.053	0.04%	0.062	0.062	0.12%
2T+Z	0.847	0.846	0.07%	0.992	0.991	0.15%
2T+2Z	3.387	3.384	0.09%	3.969	3.963	0.15%

**Table 3b:** MSE comparison for estimation at the top of a band with linear vertical trend

Combination	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
2T+0.25Z	0.084	0.069	17.81%	0.086	0.079	8.28%
2T+Z	1.141	1.032	9.54%	1.287	1.191	7.44%
2T+2Z	3.868	3.798	1.83%	4.496	4.430	1.47%

When the trend is only present in the upper side of the band, TI derived weights approaches to those obtained by Kriging using the composite variogram (see Figures 9a, 9b, 9c and 9d). At difference of the previous example, this is true even when the estimated point is located at the middle of the band (Figures 9a and 9b), where the high central weight induced by the composite variogram is closely replicated in the TI derived weights. But, whereas the Kriging weights profile are symmetrical when estimating at the middle of the band, the TI derived weights minimize the influence of the sample in the extreme of the string affected by the trend, and present a positive string effect in the extreme sample of the side where no trend is present.

Tables 4a and 4b present the Kriging and optimization MSE errors obtained from TI with different combinations of trends and residuals. In these tables it can be observed that, in general, the pronounced is the trend slope the bigger is the improvement in the MSE, and the bigger is the amplitude of residuals, the lower is the difference between the MSE obtained from kriging weights and the MSE obtained from TI derived weights.

The effect of increasing the trend slope can be explaining by the flexibility of the TI derived weights to adjust to complex trends. The effect of increasing the amplitude of residuals is explained by the increased randomness of the values; more random is a variable, more similar and uniform will be the weights derived by any method. The extreme of this is the pure nugget effect, which yields all the weights equal.

**Table 4a:** MSE comparison for estimation at the middle of a band with an upper side linear vertical trend

Combination	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
T+0.25Z	0.070	0.056	20.06%	0.077	0.067	13.70%
2T+0.25Z	0.077	0.056	27.51%	0.080	0.067	16.08%
T+Z	0.882	0.868	1.59%	1.049	1.027	2.06%
2T+Z	0.964	0.883	8.45%	1.141	1.053	7.67%
T+2Z	3.423	3.416	0.21%	4.033	4.019	0.36%
2T+2Z	3.524	3.470	1.55%	4.185	4.109	1.81%

**Table 4b:** Error comparison for estimation at the top of a band with an upper side linear vertical trend

Combination	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
T+0.25Z	0.092	0.073	20.51%	0.098	0.079	18.92%
2T+0.25Z	0.087	0.074	15.04%	0.088	0.080	8.29%
T+Z	1.064	1.009	5.18%	1.177	1.137	3.47%
2T+Z	1.330	1.112	16.33%	1.381	1.221	11.57%
T+2Z	3.692	3.671	0.56%	4.266	4.255	0.27%
2T+2Z	4.246	4.036	4.95%	4.715	4.546	3.58%

If a quadratic or cubic trend used the weights profiles obtained from the TI are similar to those obtained from Kriging when the estimation point is at the centre of the band (See Figures 10a and 10b). But, as previous cases, the TI weights approaches those obtained with Kriging using the composite variogram when the estimating a point which is not located at the middle of the domain. In general, the higher the trend order, the bigger the improvement of the MSE of the TI derived weights over the Kriging weights, thus when a cubic vertical trend is present the TI derived weights are comparatively much lower than those obtained from SK or OK.

**Table 5a:** Error comparison for estimation at the centre of a band with non linear vertical trend

Trend Type	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
Quadratic	0.934	0.862	7.71%	1.120	1.033	7.78%
Cubic	0.934	0.849	9.15%	1.242	0.999	19.58%

**Table 5b:** Error comparison for estimation at the top of a band with non linear vertical trend

Trend Type	SK	Unconstrained opt.	$\Delta E$	OK	Constrained opt.	$\Delta E$
Quadratic	1.470	1.154	21.52%	1.461	1.242	14.99%
Cubic	1.582	1.171	26.01%	1.588	1.290	18.75%

### Non Linear Estimation

The coefficients of a second order polynomial were optimized without constraints according to the procedure explained above in order to get the data conditioned weights for a vertical string with five samples. The training images used were the same as used for the first example of linear estimation; these are the TI codes p162f and p216f.

Figures 11a to 11d show the data conditioned weights obtained from the TI p162f. The curves presented there reproduce the same weight profiles as obtained using linear estimation; this is a predominance of the central sample weight for short distances evolving to a pronounced string effect at longer distances. Moreover, depending of the distance and the sample position the relation between the sample value and the sample weight can increase or decrease following a quadratic relationship.

The sum of weights (Figure 12) decreases with the distance and for this particular TI higher grade samples can receive lower weights at long distances. Compared with the SK MSE's and TI based Linear Estimation MSE's, the TI based non linear estimation MSE's are the lowest but doesn't improve the MSE differential considerably.

Data dependent weights obtained from the TI p216z also show the typical evolution towards the string effect as the distance to the estimated point increases, but also show similar arrangements to the weights obtained for linear estimation. In this particular case the high weights assigned to the upper sample at mid distances are reproduced in the data dependent weights.

As in the previous case, the MSE improvement with non linear estimation is not much higher as the obtained with linear estimation (see Figure 16).

If a perfectly Gaussian training image is used, the data dependent weights become constant for every data value (see Figure 7). This is due to the reduced connectivity between the values of different classes and the stationarity induced by the Gaussian simulation logarithm.

### Conclusions

Estimation using TI derived weights is not based on conventional assumptions of stationarity and ergodicity, the weights provide flexibility to deal with complex spatial features and relationships present in training images. The MSE error improvement using such weights instead of kriging is minimal when the Training image appears stationary with linear features. In fact, the optimized results are equal to kriging with the training image is a simulated Gaussian realization. The improvement, however, becomes significant when the TI is non stationary.

MSE improvement over kriging when using data-dependent weights in non linear estimation is not considerably greater than when using data independent weights in linear estimation. However if strong continuities of particular class values are present in the TI, this improvement can be higher.

Linear and non linear estimation based on data independent and data dependent weights from training images leads to an alternative path to deal with problematic issues in variogram based estimation techniques such as non-stationarity, unconformities, particular class values connectivity and other spatial features present in real world phenomena.

### References

- Deutsch, C.V., Kriging with Strings of Data, *Mathematical Geology*, Vol. 26, No. 5, November 1994, pp. 623-638.
- Deutsch, C.V., Kriging in a Finite Domain, *Mathematical Geology*, Vol. 25, No. 1, January 1993, pp. 41-52.
- Deutsch, C.V., Ortiz, J.M. & Bhandari, D., Optimal Weights for Linear Estimation Using Training Images. Centre for Computational Geostatistics, Report Eight, 2006.

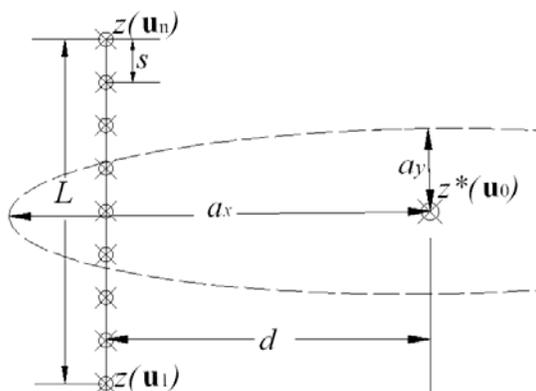


Figure 1: String data configuration

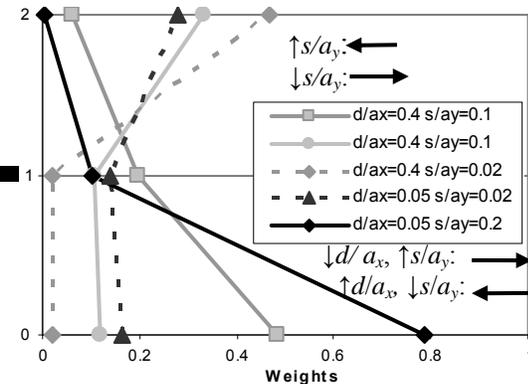
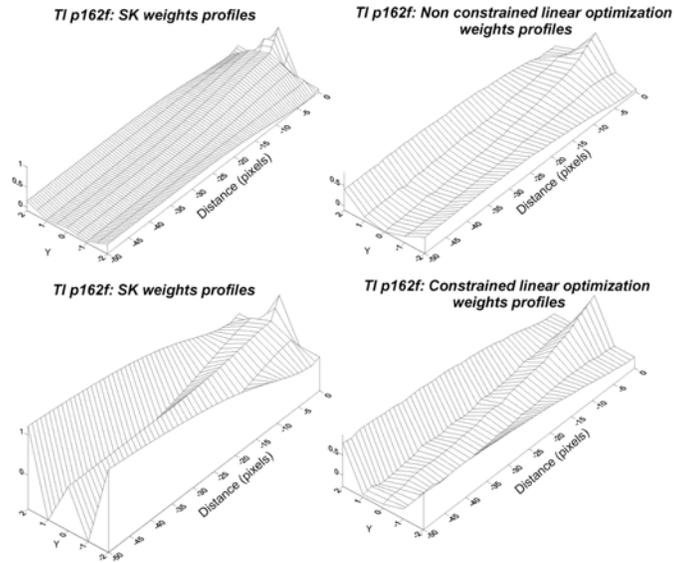
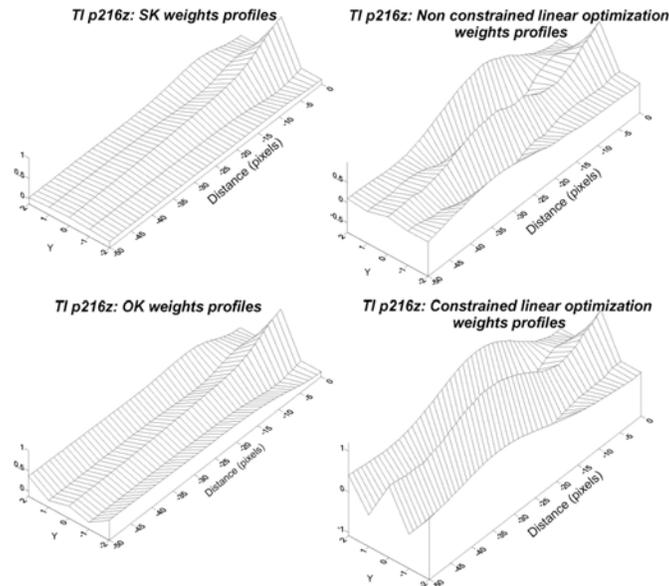


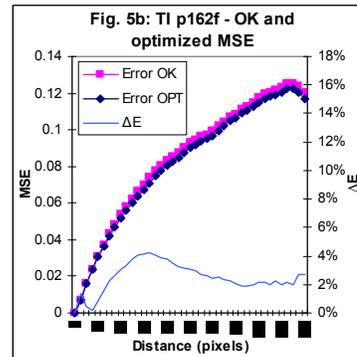
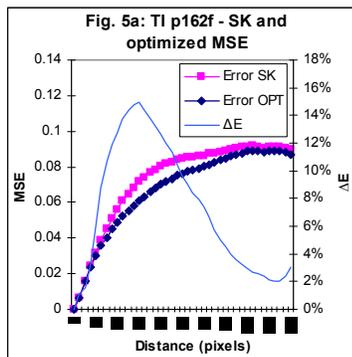
Figure 2: OK weights for the upper half of a data string with five samples

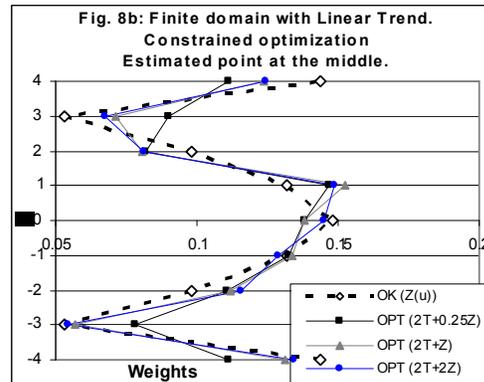
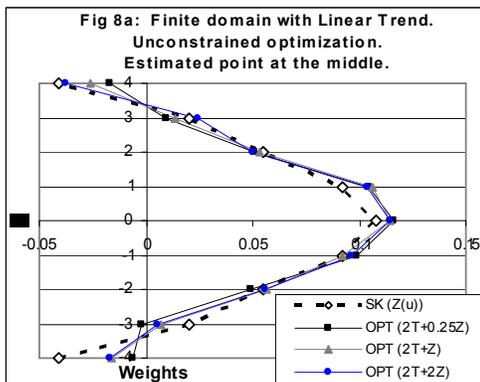
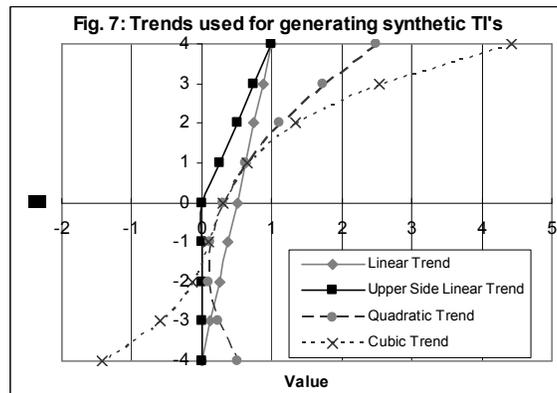
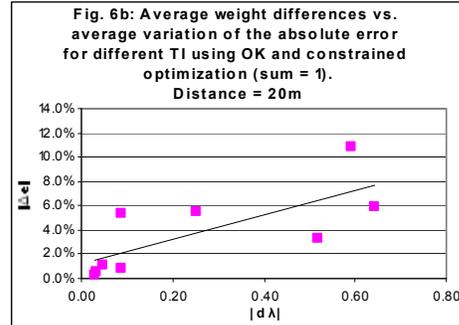
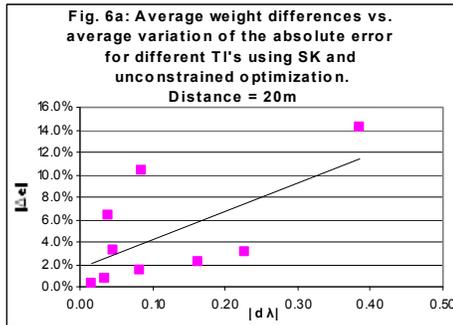
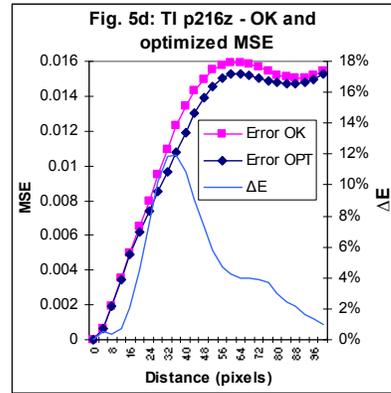
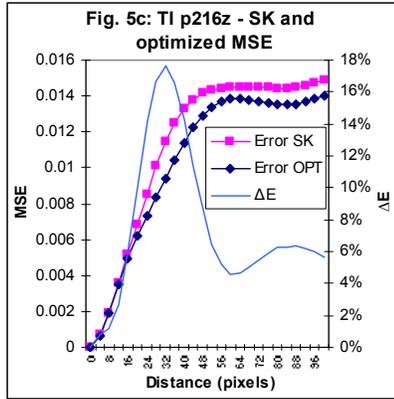


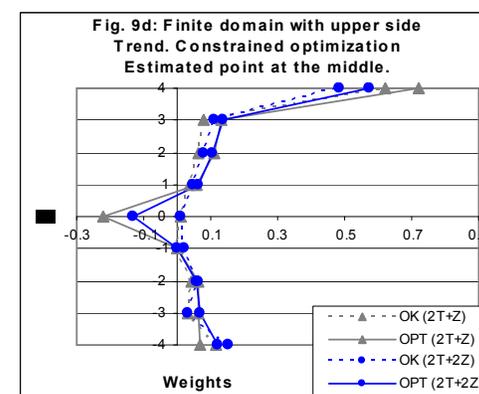
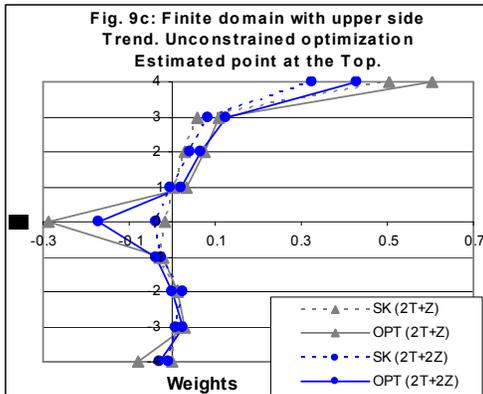
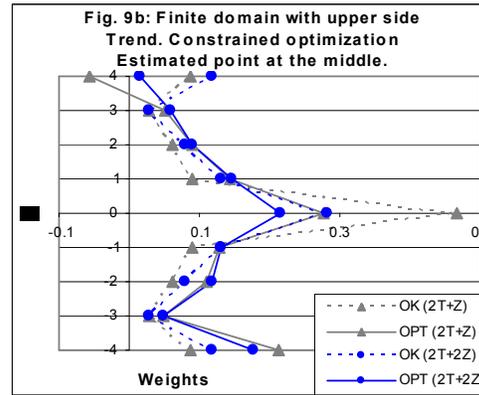
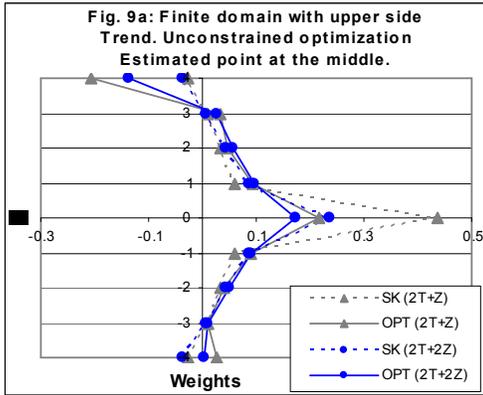
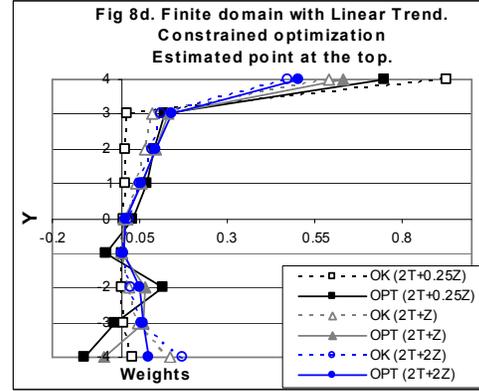
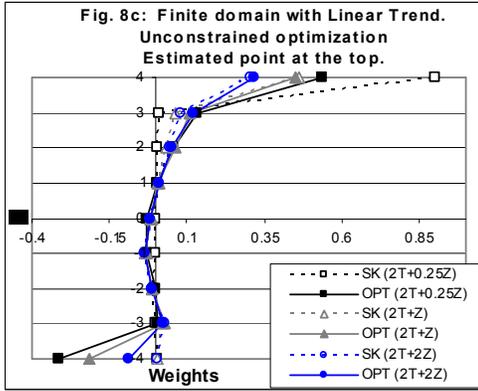
**Figure 3:** Weight profiles according the distance for Kriging and TI p162f optimization.

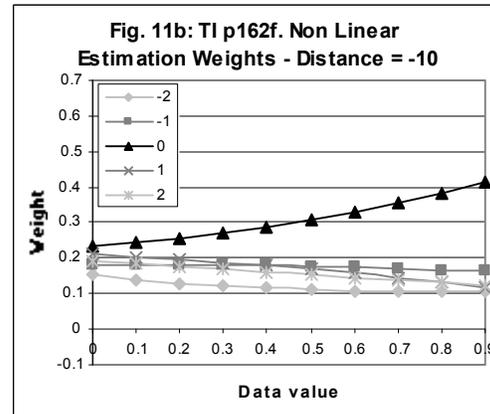
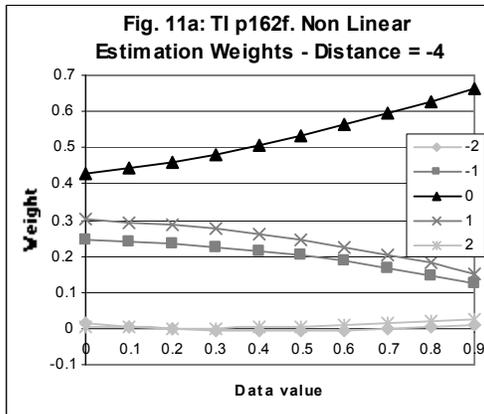
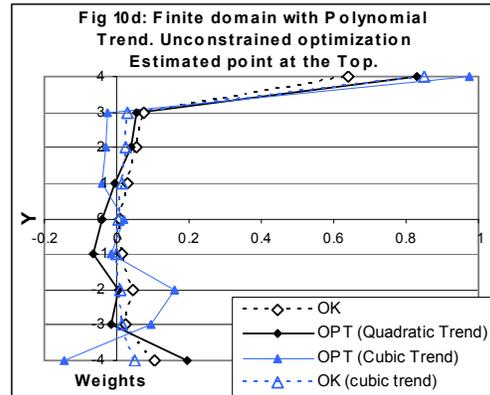
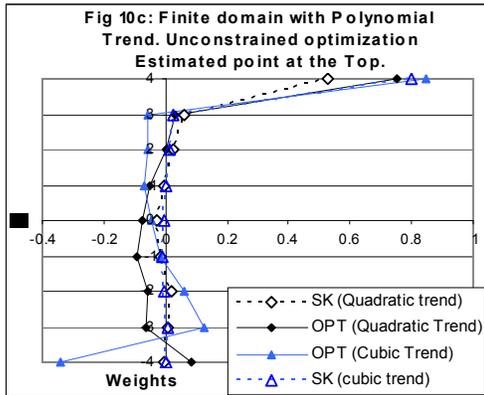
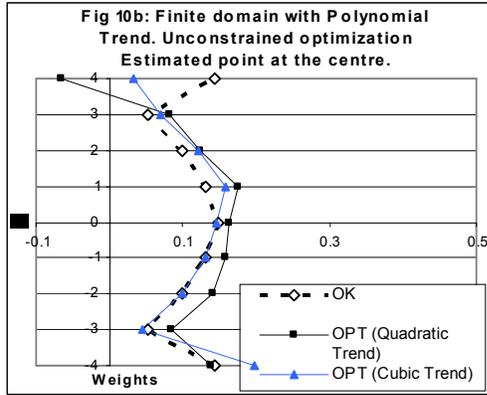
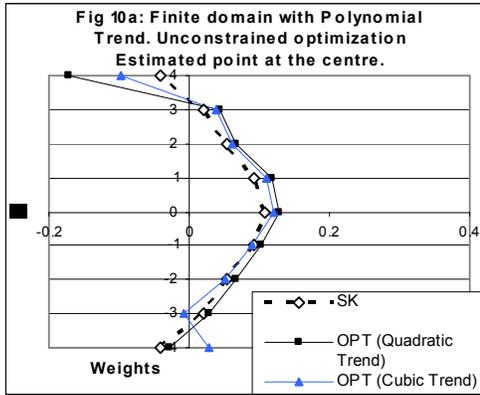


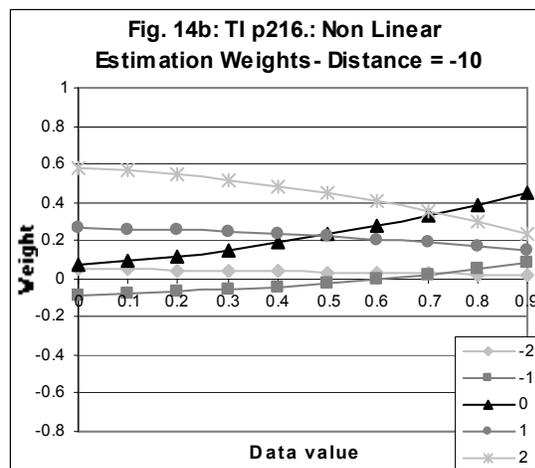
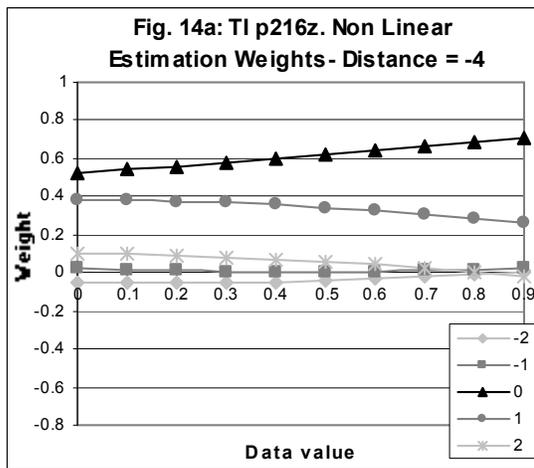
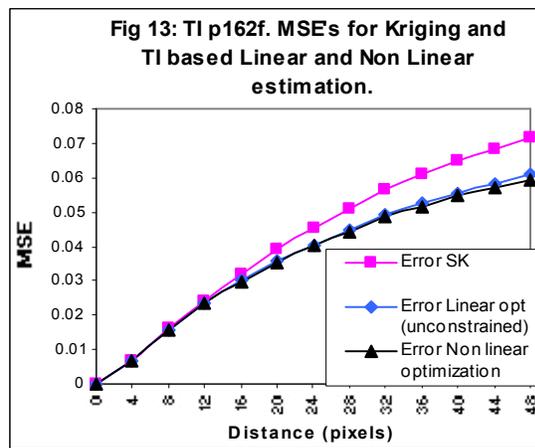
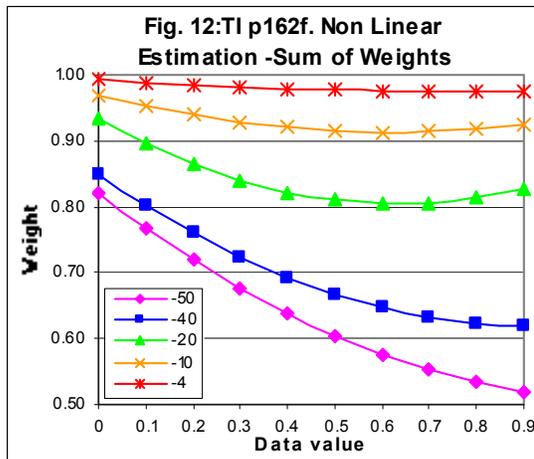
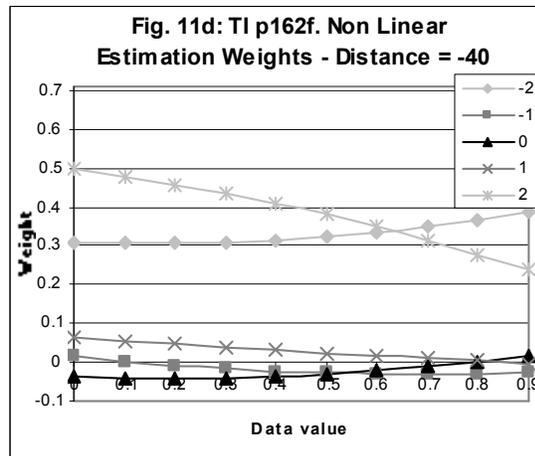
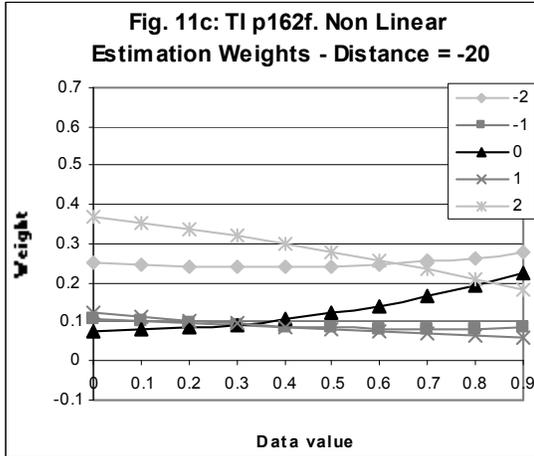
**Figure 4:** Weight profiles according the distance for Kriging and TI p216z optimization.

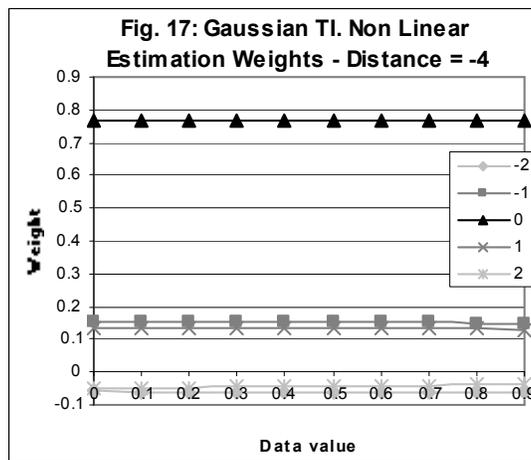
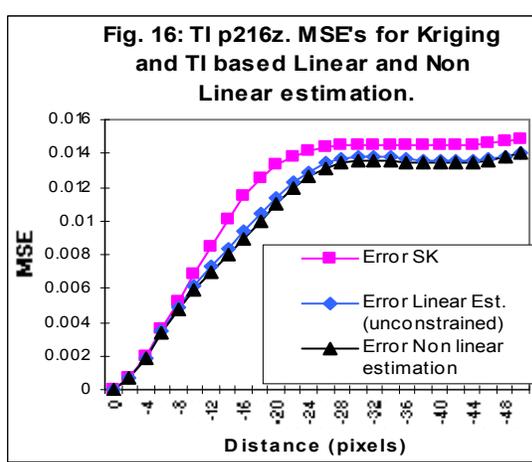
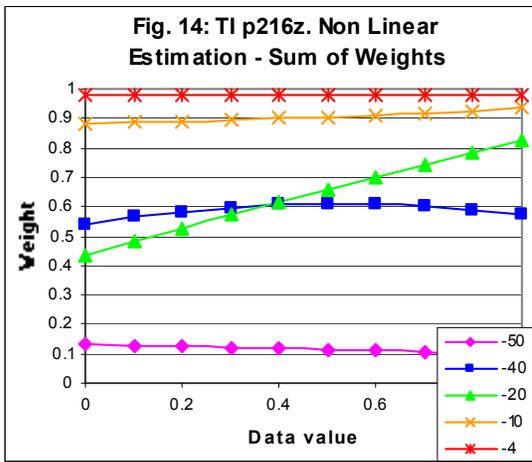
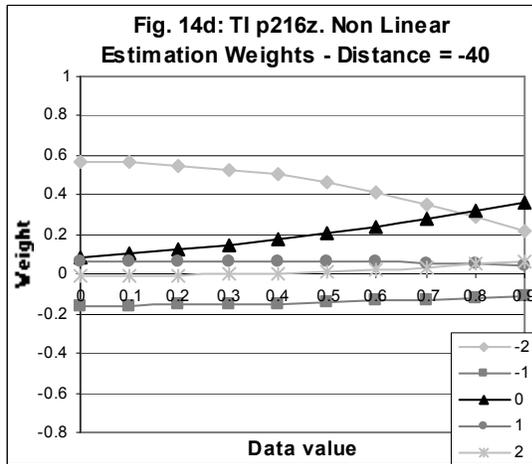
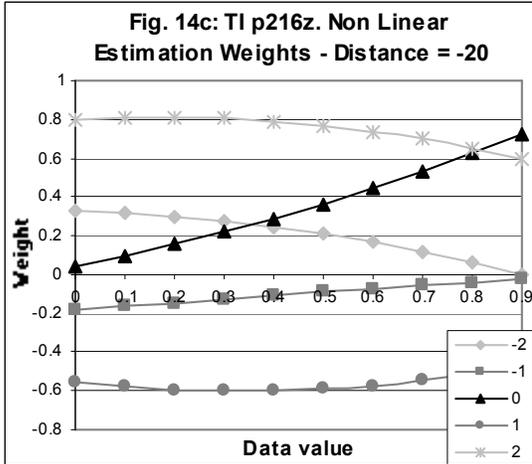












## Two of the training images used

