Modified Kriging: Evaluation, Modification, and Recommendations

C.A. Schultz
S.C. Myers

September 1997

This is an informal report intended primarily for internal or limited external distribution. The opinions and conclusions stated are those of the author and may or may not be those of the Laboratory.

Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under Contract W-7405-ENG-48.
DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (423) 576-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161
Modified Kriging: Evaluation, Modification, and Recommendations
Craig A. Schultz and Stephen C. Myers, Lawrence Livermore National Laboratory, University of California.

This report describes modifications to a technique developed by Sandia National Laboratory (SNL) to predict general corrections (traveltime, etc.) for a geographic grid only sparsely covered by calibration points (Hipp and Young, 1997). SNL has worked to create a modified version of linear prediction (kriging) based on the idea of blending the surface back to zero at some distance from the points or, in other words, by damping the solution through the damping of the input data points. LLNL has been working with SNL to evaluate Modified Kriging. This report documents our evaluation of the technique and our resulting recommendations to SNL.

Summary
Our initial evaluation of the “Modified Kriging” approach shows that the current technique tends to give a significant misprediction of the uncertainty estimate for predicted points. In addition, the final surface does not in general retain the correlation coefficient that is input to the algorithm. We have found it necessary to reformulate the “Modified Kriging” algorithm from basic principles. The basic premise for the modified kriging algorithm is the introduction of a blending (or damping) function (Hipp and Young, 1997) into the standard kriging algorithm. The blending function is used to smoothly transition the solution from a kriged estimate in areas with good data coverage to a background value and uncertainty, expressed a priori, in areas where no observations are available. Given the nature of the problem we integrate the blending function into kriging using Bayesian methodology. By introducing the blending function in a Bayesian framework, the data and damping statistics are implemented consistently throughout the kriging matrices, allowing proper propagation of errors and an accurate estimation of uncertainty for the estimated values. This technique maintains the desirable attributes introduced by the blending function while retaining the statistical benefits of ordinary kriging.

Modified Kriging: Constraints on the Solution
Kriging is a common technique to estimate a spatially dependent variable given some distribution of observations (Srivastava et al, 1989 - Oxford Press), with the predicted value calculated by a weighted sum of the observed data values. Simple kriging utilizes measured values and a variogram (related to the true spatial covariance of the observations) to generate an un-biased prediction of the spatial variable. In seismology, distributed correction values can be thought of as a spatially dependent variable with an estimated correlation length, prompting the use of kriging to predict residuals at unsampled localities. However, seismic data sets tend to be heavily sampled in some areas, while other areas are devoid of samples. The special character of seismic data sets requires some modifications of the kriging technique. Hipp and Young (1997; SNL) have developed a promising ad hoc modification to ordinary kriging that takes into consideration seismic data sets by incorporating the following constraints:

1. Incorporation of differing certainties (sample variance) for observations allows the introduction of different grades of ground truth.
2. Smoothness and continuity of the surface both locally and globally required for effective implementation in EvLoc
3. Predicted variance blends smoothly and continuously to a regional background incorporation of an a priori background model distribution
4. Reduce (damp) the influence of uncertain data when predicting unsampled points well behaved solution in absence of data points and allow for additional selective downweighting of data points
5. Provide smooth and continuous residual/error surface transitions at regional boundaries allow for juxtaposition of different regional correction surfaces

SNL accommodates the above requirements by tapering the magnitude of each observation as a function of distance from the observed point. The distance-dependent reduction in magnitude is parameterized with an analytic “blending” function, which takes on the observed data value at the location of the measurement and smoothly decreases to zero (or some background value) at a specified distance. A weighted sum of the blended (or damped) values and variances are used to estimate the value and uncertainty at unknown points. Use of “blended” data values has the effect of damping the predicted value to an a priori background value in areas where no observations are available. The predicted value, with damping, can be expressed as (2.1):

$$Z_p^*(\bar{X}_p) = \sum_{i=1}^{N} \omega_{ip} Z_i \left( \left| \bar{X}_i - \bar{X}_p \right| r_i \right)$$

where $Z_i(h)$ represents the actual correction value multiplied by a unit normalized blending that damps the assigned value as a function of horizontal distance, $h$, from that point. In this formalism the damping can be different for each data point, allowing adjustment of the damping for each data point. It is for this reason that in this short report we freely interchange the terms “blending” and “damping” function. Given (2.1) as the adjusted form for the linear predictor, the estimated uncertainty of the predicted value can be written as (2.2),

$$\sigma^2_p(\bar{X}_p) = E \left\{ \sum_{i=1}^{N} \omega_{ip} \left( Z_p - Z_{ip} \right)^2 \right\} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \omega_{ip} \omega_{jp} \left( Z_{ip} - Z_{jp} \right)^2$$

where the parenthetical brackets have been removed for convenience and $Z_p$ represents the unknown value at the predicted point. We can express the variance between points $Z_i$ and $Z_j$ as (2.3),

$$E \left\{ \left( Z_i - Z_j \right)^2 \right\} = 2\gamma(Z_i, Z_j) = C_z(Z_i, Z_i) + C_z(Z_j, Z_j) - 2 C_z(Z_i, Z_j)$$

where $C_z$ represents the nonstationary covariance between two points and $\gamma$ is the nonstationary variogram function. By minimizing equation (2.2), the final weights and the predicted value are determined. To this point we have followed the original kriging
derivation and this equates to the original SNL formulation. It is at this point that we turn to a Bayesian approach to guarantee that statistical uncertainties are properly propagated through the system.

**Justification of further modification**

Our effort to benchmark the original technique shows that the current technique tends to give a significant misprediction of the uncertainty estimate for predicted points. The approach that we use for benchmarking is as follows:

- Generate surface with known statistics
- Randomly sample this surface
- Enter known covariance and uncertainty as 'a priori' knowledge
- Estimate correction value and uncertainty with modified kriging
- Determine accuracy of estimates by direct comparison
- Determine accuracy of uncertainty estimate by its statistical distribution normalized to a standard Gaussian

Predictions giving background or zero uncertainty were excluded from the final statistics. Thus, only predicted values that are within the zone of influence of the data are tested.

![Figure 1: Comparison of Bayesian Modified Kriging and current Modified Kriging uncertainties for a known surface covariance and a blending function with a 3\(^\circ\) range. The current Modified Kriging approach gives a significant misprediction of the uncertainty estimate and most comparisons (shown in detail later in the paper) demonstrate an even greater degradation of the uncertainty estimate.](image)

Figure 1 illustrates one of the better matches between the two techniques found using the above approach. However, even in this case, there tends to be a significant underprediction of the uncertainties for the predicted value. In addition, the final surface does not in general retain the correlation coefficient that is input to the algorithm. We have found it necessary to reformulate the "Modified Kriging" algorithm from basic principles.

**Modifications to the original approach**

After examining the original Modified Kriging technique, we suggest that additional criteria be added to those listed in 1-5:

6. Correlation function has primary effect on final predicted surface correction statistics
the statistics of the final correction surface needs to reflect our a priori statistics and not be controlled entirely by the blending/damping function

7. Reversion to simple (optimal) kriging in regions where data coverage is good.

8. Damping can only act to reduce the correlation function, not extend it locally.
   damping the data should only act to decrease our maximum estimate of the zone of influence of a data point on the correction surface

9. Variable damping that increases when data coverage is poor and values are extrapolated.
   need increased damping when we approach poor data coverage to stop spurious extrapolation

10. Estimated uncertainties need to properly account for 6) through 8) above.
    imperative to location that the final uncertainty estimate be accurate, this needs to be benchmarked

An approach that differs somewhat from the original approach may be used to accurately incorporate constraints 6-10, while still abiding by the original constraints 1-5. In order to satisfy these constraints, while maintaining statistical rigor, the effect of the damping function must be properly incorporated into the covariance equations. Eq. (2.3) can be viewed as the covariance of the original data surface that is adjusted by an additional damping function, following (3.1).

\[ C_Z(Z_i, Z_j) = C_Z(Z^0_i, Z^0_j) \]

where \( Z^0_i \) is the actual value of a data point multiplied by the normalized damping (blending) function. This covariance is no longer just the covariance of the data points, but the covariance of the data points with an additional damping function. In a Bayesian sense (Omre, 1987 - Mathematical Geology), this damping may be accounted for 'a priori' by assuming an 'a priori' variogram model of the form (3.2),

\[
\gamma_M(Z_i, Z_j) = \frac{1}{2} \rho^M_{ij} (\sigma_i, \sigma_j - \hat{Z}_i \hat{Z}_j \sigma_i, \sigma_j) = \gamma_{Z,M}(Z_i, Z_j) - \gamma_{ZM}(Z_i, Z_j)
\]

where \( \sigma^2 \) corresponds to the variance at a point along the surface. The second equality in (3.2) simply states that the variance equates to the joint probability of damping the original variogram minus the conditional probability of the original variogram. This 'a priori' model variance accounts for the fact that we are increasing our uncertainty of the predicted point in regions where the predicted value is being damped to zero. As a result, it can only act to reduce the 'a posteriori' correlation between points in the model. By its very nature, damping (blending) acts only to decrease the correlation length of the surface and this is reflected in this function.

The final system of kriging equations that combines the 'a priori' knowledge about the surface with 'a priori' knowledge about the damping factor for each data point can be expressed as (3.3),

\[
\sum_j \omega_j [\gamma^y_{Z,M} + \gamma^y_M] + \lambda = [\gamma^y_{Z,M} + \gamma^y_M]
\]
where \( \lambda \) is the Lagrange parameter that allows for the introduction of the unbiasedness constraint. Written in the notation of the original SNL report this is (3.4),

\[
\sum_{j=1}^{N} \omega_{jp} \left[ \sigma_{ip}^2 + \sigma_{jp}^2 - 2 \rho_{jp} \sigma_{ip} \sigma_{jp} \right] + \lambda \left[ \sigma_{ip}^2 + \sigma_{pp}^2 - 2 \rho_{ip} \sigma_{ip} \sigma_{pp} \right] = 1
\]

Finally, the estimated uncertainty in the predicted value can be expressed as (3.5),

\[
\sigma_{ip}^2 = \frac{1}{2} \left[ \lambda + \sum_{n=1}^{N} \omega_{ip} \left[ \sigma_{ip}^2 + \sigma_{pp}^2 - 2 \rho_{ip} \sigma_{ip} \sigma_{pp} \right] \right]
\]

which as we will show gives an accurate prediction of the errors when the surface statistics are approximately known.

Attributes of Bayesian Kriging
The ‘a priori’ model variance assumed above guarantees that constraints 1 through 10 listed in the first section are met. This “Bayesian Kriging” analysis properly accounts for the predicted uncertainties arising from the damping constraint. Performing a similar Bayesian analysis on the original Modified Kriging formulation shows that the damping function almost completely washes out the correlation functions influence. This dominance of the blending function (BL) is not a desirable effect, because the statistics (and the error estimates) are held primarily in the correlation function. For example, if the “range” of the BL is large, then the covariance between distant points was estimated to be large, in spite of the designated correlation function. The artificially large covariance results in error estimates that are significantly too small. In summary, Bayesian Kriging allows application of the BL to dampen the input statistics and data values towards an a priori background model in unsampled regions, while allowing data values and statistics to be used when estimating points in sampled regions.

Examples of the Bayesian Modified Kriging Approach
This section gives some examples of surfaces estimated with the original Modified Kriging (MK) and Bayesian MK (LLNL) approaches. We present examples using both a synthetic surface with known spatial. In these examples the MK and Bayesian MK routines give good estimates of unsampled points, but the Bayesian approach gives more appropriate error estimates.

Figure 2 is a synthetic surface with zero mean, variance of 4, and a known spatial correlation function. In order to test the kriging routines, we randomly sampled this surface at 100 points, then input the 100 samples into each of the kriging routines to recover the surface. The known correlation function was input into the kriging routines, giving a best-case test. We tested three, end-member cases:

1. BL range of one half the correlation distance (range=3; moderately damped);
2. BL range of three times the correlation distance (range=18; loosely damped);
3. BL range essentially infinite (no damping).

Figures 3-5 shows the recovered surfaces using BLs with ranges of 3, 18 and infinity, respectively. Figure 3 shows little difference in the original and Bayesian MK results; because, the correlation length is dominated by the BL in both instances. However, as the range of the BL increases (Figure 4), results from the two methods diverge. At a BL range of 18, as shown in Figure 5, the original MK gives inappropriately small uncertainties. As the range of the BL increases, the original MK error estimates continue to shrink. The discrepancy between estimated and actual error is illustrated at the bottom of Figures 3-5. This shows the theoretical (curve) and observed (histogram) error distribution normalized to the estimated variance. For the original MK the observed error distribution is significantly broader than predicted, with a high percentage of residuals falling outside the 95% confidence interval. The Bayesian methodology gives a far better estimate of the uncertainty.

The uncertainty in the original MK is inappropriately small (for large BL range), because the correlation length is controlled by the BL range, regardless of the specified correlation range. Thus for infinite BL range, all of the observations are assumed to be perfectly correlated with each other and with the unknown points, resulting in perfect certainty at the estimated point. In the Bayesian formulation, the BL can only decrease the correlation between two points. For Bayesian MK, if the correlation distance is shorter than the BL range, then the correlation distance controls the covariance matrix. However, if the BL range is less than the correlation distance, then the BL range reduces (dampens) the covariance between any two points (which forces the uncertainty in the estimated point towards the a priori uncertainty). Blending (damping) between kriged and a priori values and uncertainties are consistent in the Bayesian approach.

Conclusion
The introduction of blending functions into kriging allows great flexibility in damping the solution to a priori values and uncertainty. This is a powerful modification to the ordinary kriging approach. When formulated in a Bayesian framework, the blending function can be introduced into the kriging matrices in a statistically consistent manor, allowing proper estimates of uncertainty. Based on the above analysis we highly recommend this Bayesian Modified Kriging for implementation. Other recommendations include:

- the use of only positive-definite correlation functions due to inherent instabilities in general kriging approaches. A check for positive-definiteness should be added into the code.
- an additional adjustment to Modified Kriging to account for spurious predicted values at the edge of data sets where the surface is extrapolated. By introducing a measure of data coverage similar to the location coverage measure (i.e. maximum azimuthal gap) one can implement a nonstationary taper that is a function of this gap. As the Maximum gap approach 180°, the damping function can be changed to stronger damping. This would allow for a surface that is well-behaved in areas of poor coverage, but allows for a removal of damping and a return to standard kriging in regions where data coverage is good. Such an approach might be termed “Threshold Bayesian Kriging”.

Figure 2: Correction surface generated synthetically using a fractal covariance function. The correlation function and background uncertainty of this surface is known exactly. This surface is used to rigorously benchmark the Modified Kriging and Bayesian Modified Kriging algorithms. This benchmarking involves taking random samples from this surface and comparing predicted values and errors with the actual surface. The accuracy of the uncertainty estimates can then be rigorously evaluated.
Figure 3: Comparison between Bayesian Modified Kriging and the original Modified Kriging Routine. The estimated correction surface (a) is shown for both cases along with the location of the profiles taken across the surface. These profiles (b) are shown along with the estimated uncertainty bars and the profile of the original surface that was sampled. The final plot (c) shows how well the observed error (histogram) matches the predicted error (curve). Bayesian Modified Kriging appears to give a better uncertainty
Figure 4: Similar to figure above, but blending function now has a range of 18°. The Bayesian Modified Kriging approach produces a good estimate uncertainty estimates that are consistent with the observed errors.
Figure 5: Similar to figure above, but blending function now has a range of 1000. The Bayesian Modified Kriging approach produces a good estimate uncertainty estimates that are consistent with the observed errors.