A Short Note on:

Integration of Multiple Secondary Data in Geostatistical Reservoir Modeling

Clayton V. Deutsch

Increasingly there are multiple secondary data variables to be used in geostatistical reservoir modeling: acoustic and elastic impedance, other seismic attributes, production data, and geological trend maps. There is a need to simultaneously account for these variables. Robust techniques are needed to account for multiple secondary data variables.

One evident alternative to direct accounting for two or more secondary data is to combine the secondary data and use well-known techniques to account for one secondary data, that is, construct a Y variable as a function of the multiple data, $Y=f(Y_1, Y_2, \ldots)$. Regression or neural networks could be considered for this purpose. This would appear to be a good option when the secondary data arise from the same underlying data source, as in seismic. This would not likely work as well when the secondary data arise from data at different scales and/or measuring different underlying physical properties.

Block Cokriging – the Correct Approach

Leaving aside lesser known iterative techniques, the correct theoretical approach to account for multiple secondary data is block cokriging. The simulation of the primary Z variable proceeds sequentially where a kriging is performed at each grid node. The kriging estimator for two secondary data variables takes the following form:

$$z^*(\mathbf{u}) = \sum_{\alpha=1}^{n_z} \lambda_\alpha z(\mathbf{u}_\alpha) + \sum_{\alpha=1}^{n_{y_1}} \lambda'_\alpha y_1(\mathbf{u}'_\alpha) + \sum_{\alpha=1}^{n_{y_2}} \lambda''_\alpha y_2(\mathbf{u}''_\alpha)$$

where there are three sources of data, hence three summations over n_z, n_{y1}, and n_{y2}, which are the number of primary and secondary data. The λ's are the kriging weights. In general there are no restrictions regarding the units of the data, the volume support of the data, the size of the block being kriged/simulated, or the number of data. In practice, there are computer and inference limitations that must be considered. The (n_z x n_{y1} x n_{y2}) by (n_z x n_{y1} x n_{y2}) matrix set of equations that must be solved can be partitioned into 3(by 3) regions.

$$\begin{bmatrix} \overline{C}_{Z,Z} & \overline{C}_{Z,Y_1} & \overline{C}_{Z,Y_2} \\ & \overline{C}_{Y_1,Y_1} & \overline{C}_{Y_1,Y_2} \\ & & \overline{C}_{Y_2,Y_2} \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda' \\ \lambda'' \end{bmatrix} = \begin{bmatrix} \overline{C}_{Z,Z} \\ \overline{C}_{Z,Y_1} \\ \overline{C}_{Z,Y_2} \end{bmatrix}$$

Each term in the matrix equation above is a series of volume averaged covariance, for example, $\overline{C}_{Z,Y_1}$ represents volume averages of a Z data location and a Y_1 block.

Some limitations of this "correct" approach include: (1) relatively complex software with no commercial alternatives, (2) CPU-expensive calculation of point-block and block-block (cross) covariances, and (3) the difficulty of establishing the correct histogram at the correct scale.

Bayesian Updating

The Bayesian updating formalism is particularly useful in presence of one secondary variable. In fact, we can show that it is identical to collocated cokriging. This equivalence between Bayesian updating and collocated cokriging does not hold for more than one secondary data variable. The basic idea of Bayesian updating may be written:

$$p_k^{**} = C \cdot p_k^{*} \cdot \frac{p_k^{'}}{p_k} \cdot \frac{p_k^{''}}{p_k}$$

where the $p^{**}{}_k$'s are the final probabilities, C is a normalization constant, the $p^{*}{}_k$'s are the probabilities from indicator kriging with hard data alone, the $p^{'}{}_k$'s are the probabilities from first secondary data, the $p^{''}{}_k$'s are the probabilities from second secondary data, and the p_k's are the global proportions or probabilities.

The main limitation of this formalism is that the multiple secondary data must be conditionally independent. If they are not too much weight is given to the secondary data. We could think about weighting these terms:

$$\hat{p}_k = C \cdot p_k^{*} \cdot \left(\frac{p_k^{'}}{p_k} \right)^{w_1} \cdot \left(\frac{p_k^{''}}{p_k} \right)^{w_2}$$

Where w_1 and w_2 are equal to 1 for independent secondary data and equal to ½ for redundant secondary data. One could assume that this exponent follows a linear relationship with the correlation between the secondary data, see Figure 1. Of course, rigorously accounting for redundancy between the secondary data would require some investigation.

Collocated Cokriging

Collocated cokriging is a remarkably simple and effective approach to integrate secondary data. The main advantage of collocated cokriging over Bayesian updating is that redundancy between the secondary data is accounted for in the kriging formalism.

Needed covariance functions may be estimated by the typical Markov approximation, which sets the covariance function as proportional to the *Z*-covariance:

$$C_{Z,Y_1}(\mathbf{h}) = \rho_{Z,Y_1} \cdot C_{Z,Z}(\mathbf{h})$$

Where $C_{Z,Y1}(\mathbf{h})$ is the cross covariance between Z and Y_1, $\rho_{Z,Y1}$ is the correlation between collocated Z and Y_1 values, and $C_{Z,Z}(\mathbf{h})$ is the covariance of the primary Z data values. There must be some constraint on the correlation coefficients $\rho_{Z,Z}$, $\rho_{Z,Y1}$, $\rho_{Z,Y2}$, and $\rho_{Y1,Y2}$ to ensure positive definiteness.

Limitations include (1) the implicit assumption that the primary data, grid nodes being simulated, and all secondary data are at the same scale, (2) no use of cross spatial relationships, for example, an isolated high secondary variable is treated identical to a high value in a large region of high values, and (3) the Markov model does not permit any accounting for different spatial continuity of the secondary data from the primary data.

Other Alternatives

Kriging with an external drift may be extended to multiple secondary data variables, but the lack of explicit control over the correlation between the primary and secondary data is a cause of concern. IRF-k (intrinsic random functions or order k) could also be applied to this problem, but the difficulty of obtaining a licit model of coregionalization would force us to use the automatic

fitting. The consequences of this are (cross)variograms with too-high nugget effects and unrealistic heterogeneity in the final models.

Special Case of Collocated Cokriging for Merging Secondary Data

As stated above, one alternative to direct accounting for two or more secondary data is to combine the secondary data and use well-known techniques to account for one secondary data, that is, construct a *Y* variable as a function of the multiple data, $Y=f(Y_1, Y_2, \ldots)$. Regression or neural networks could be considered for this purpose. Simple collocated cokriging could also be used:

$$y^*(\mathbf{u}) - m = \lambda' \cdot (y_1(\mathbf{u}) - m) + \lambda'' \cdot (y_2(\mathbf{u}) - m)$$

where the variable *y* could be porosity, lithofacies proportions, or any other petrophysical property, *m* is the mean, **u** is the location being considered, λ' is the weight to variable 1 and λ'' is the weight to variable 2. This could also be written as:

$$y^*(\mathbf{u}) = \lambda' \cdot y_1(\mathbf{u}) + \lambda'' \cdot y_2(\mathbf{u}) + (1 - \lambda' - \lambda'') \cdot m$$

The cokriging equation to solve for the two unknown weights:

$$\begin{bmatrix} C_{Y_1,Y_1} & C_{Y_1,Y_2} \\ C_{Y_2,Y_1} & C_{Y_2,Y_2} \end{bmatrix} \begin{bmatrix} \lambda' \\ \lambda'' \end{bmatrix} = \begin{bmatrix} C_{Y,Y_1} \\ C_{Y,Y_2} \end{bmatrix}$$

A special case is when the two secondary data are uncorrelated, that is, $C_{Y_1,Y_2} = C_{Y_2,Y_1} = 0$, in which case the weights are equal to the correlation coefficient between the secondary variable and the variable being estimated, e.g., for variable one $\lambda' = C_{Y,Y_1} / C_{Y_1,Y_1} = \rho_{Y,Y_1}$. In general, however, the weights are less due to correlation between the secondary data. There is almost always such correlation if only because the secondary data are measuring the same underlying petrophysical properties. We can solve this 2 by 2 set of equations to get the weights. The results for standardized variables (so we use correlation coefficients):

$$\lambda' = \frac{\rho_{Y,Y_1} - \rho_{Y,Y_2} \cdot \rho_{Y_1,Y_2}}{1 - (\rho_{Y_1,Y_2})^2} \text{ and } \lambda'' = \frac{\rho_{Y,Y_2} - \rho_{Y,Y_1} \cdot \rho_{Y_1,Y_2}}{1 - (\rho_{Y_1,Y_2})^2}$$

The weights are undefined (matrix is singular) when the correlation between the two variables is 1 or –1, which is understandable because the variables are perfectly redundant with each other; however, we see that at the limit the weights approach ½ of the correlation between the secondary and primary variables. See Figure 2 below for an example. In this case the correlation between both secondary variables and the primary variable is 0.6; the weight to variable 1 starts at 0.6 when the two secondary are uncorrelated and decreases to 0.3 for perfect correlation. Of course, when the two variables are perfectly correlated they are redundant and the same and 0.3+0.3=0.6 so cokriging works.

This approach permits merging multiple secondary data to simplify geostatistical simulation.

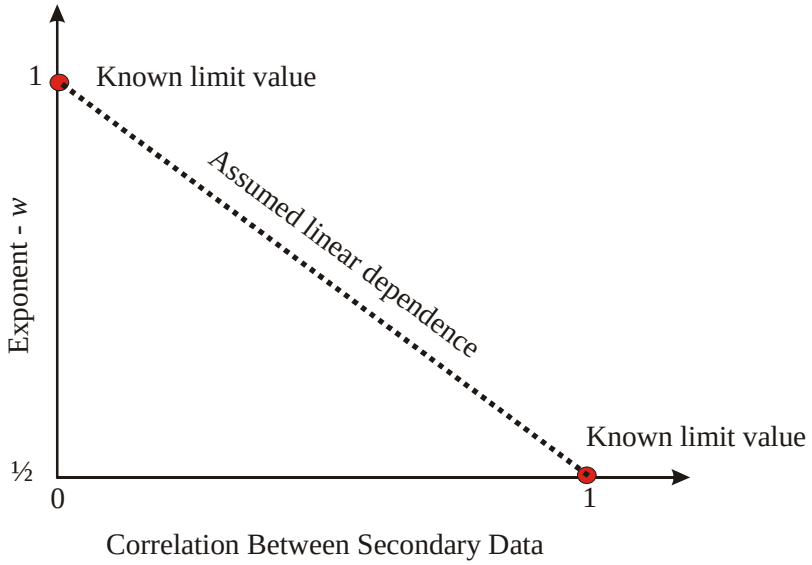

Figure 1: Dependence of exponent used to correct Bayes relation using multiple secondary data variables. The exponent must decrease as the correlation between the two secondary variables decreases. The lower limit is 1/n, where n is the number of variables: 2 in this case.

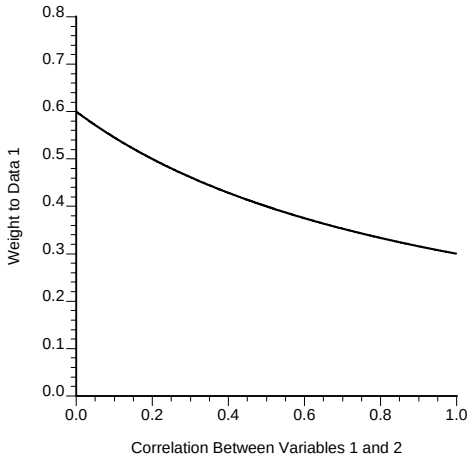

Figure 2: Dependence of the weight to a secondary data in collocated cokriging when the correlation between each covariate and the variable of interest in 0.6. The lower limit is 1/n multiplied by 0.6, where n is the number of variables: 2 in this case.