A Short Note: When do Multiscale Crossplots Make Sense?

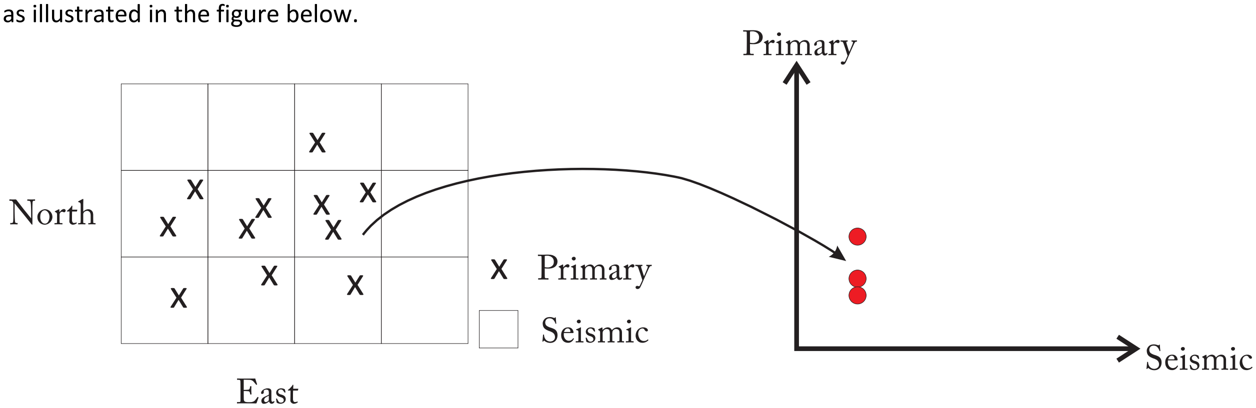
Sahyun Hong and Clayton V. Deutsch

This short note discusses multiscale crossplots, that is, a plot of two variables measured at different scales at approximately the same location. For example, consider high resolution well log data and low resolution seismic seismic data. Many geostatistical techniques work with bivariate or multivariate relationships between variables, however, most do not explicitly address the difference in scale. Data are often assumed to be of the same scale. This note describes the issue of multiscale multivariate modeling and summarizes how the relation modeling can be applied in data integration.

1. Bivariate Modeling(crossplots) of Multiscale Data

Cross plots of two variables is quite commonly seen in geostatistical modeling and bivariate or multivariate relation is modeled based on the cross plots. Those cross plots can be plotted when two variables are homotopically sampled and we assume, despite not explicitly specifying, that two variables have the same scale. There is no theoretical limitation that cross plot of two variables with different scales also can be plotted. Besides, some previous works implicitly assumed that two variables that have different scales are bivariate Gaussian (Lee et al., 2002; Efendiev et al., 2008).

In the case of multiscale multivariate modeling, one variable in small scale is deemed to be collocated within another variable in large scale. For example, many samples of the primary data can be included in a larger seismic scale and they are plotted on the cross plot having different primary data value with a unique seismic value as illustrated in the figure below.



The experimental plot, therefore, is a cross plot of two variables with different scales. Summary statistics calculated from usual cross plots can be calculated: linear/rank correlation coefficient, regression line and conditional expectation. The bivariate distribution can be nonparametrically modeled based on the crossplots. Once the bivariate relation is modeled, conditional mean and variance of the primary variable can be derived given the large scale secondary variable.

3. Examples

The theoretical validity of multiscale crossplots is not fully investigated in this note. However, small example shows that the approach is reasonable in practice. Figure 1 is a result of multiscale bivariate modeling and the derived primary estimate from the bivariate model. The large scale secondary data has 5 times large resolution in the vertical direction than the modeling or primary data scale as shown in the top right in Figure 1. Thus, the estimates of the primary variable are assigned at every 5 cells within the large scale. The modeled bivariate distribution is shown in the middle. The kernel density estimation is used and some marginal constraints are checked and honored (see CCG paper 405 in this volume). The downscaled primary estimates are shown in the bottom. Only three Monte Carlo simulation results are plotted. P-values for Monte Carlo simulation are spatially correlated values obtained from the primary variable variogram. The average of simulations reproduces the input large scale secondary data because the primary estimates are conditioned to the secondary data values.

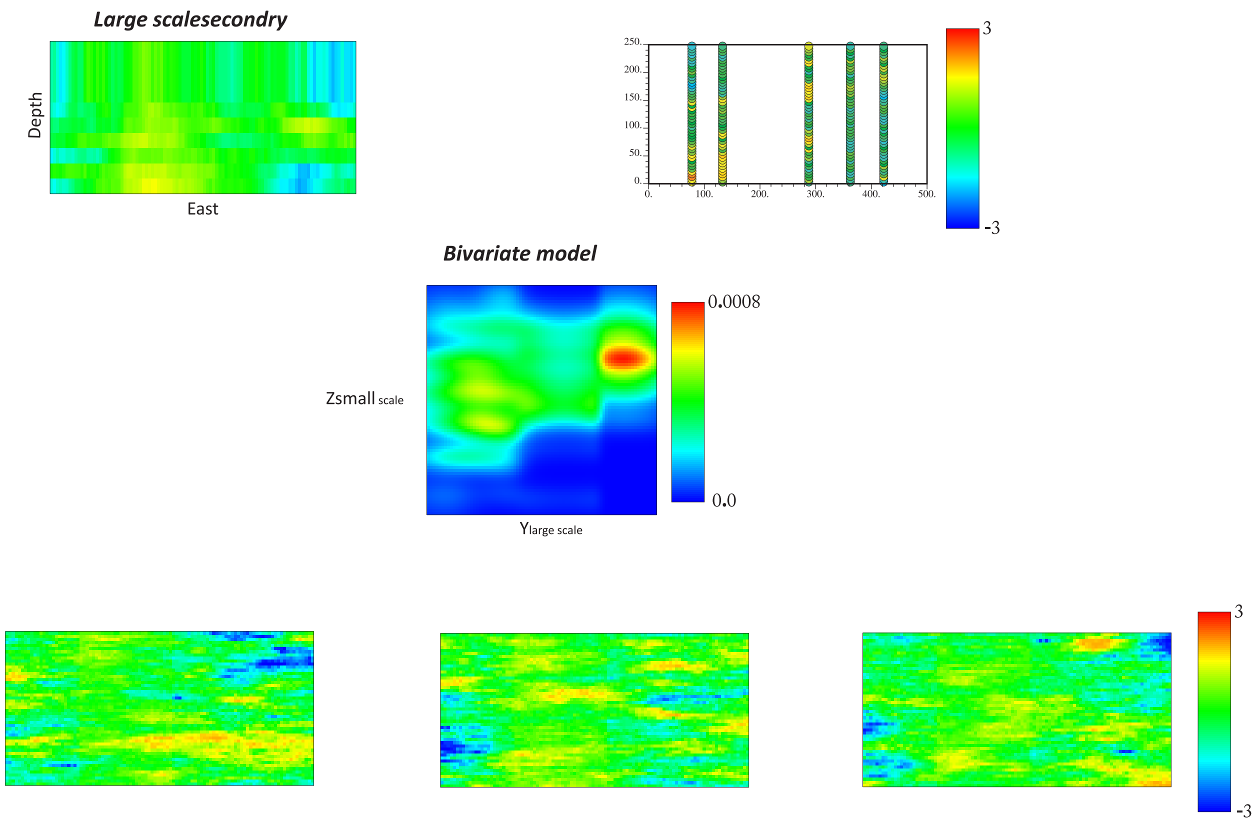


Figure 1: Example of the multiscale bivariate modeling.

4. Considerations

The multiscale crossplots may be reasonable if the scale difference between two or more variables is not significant. For example, case (A) shown in the figure below may not be a desirable case for the multiscale bivariate modeling rather than case (B). Randomly assigning the estimated primary variable within the large scale secondary block will ignore the spatial variability of the primary variable.

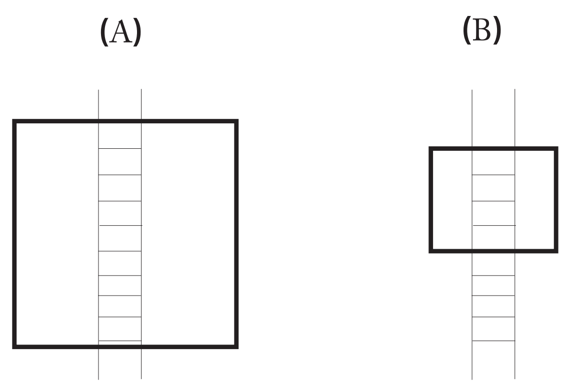


Figure 2: Scale difference between primary and secondary variable. Case (B) would be a suitable case for the application of multiscale bivariate modeling.

Another consideration in the cross plot of multiscale data is the relative position of the small scale values. A small scale sample in the center of the larger volume should have a higher correlation with the data at the large scale than a sample near the outside. This could be computed theoretically in some cases (see Oz and Deutsch, 2002).

3. Potential Applicability to Bayesian Updating

The mean and variance of the primary variable are derived conditioned to the large scale secondary data. Thus, the estimated moments are at the secondary data scale. The primary data of course are variable within the secondary data scale. To account for the variability of the primary variable within a secondary data block, a Monte Carlo simulation is applied to the extracted conditional distribution given the secondary data. The simulated primary values (given the secondary data) are randomly assigned within a secondary variable scale. By doing so, there are multiple realizations of the primary variable conditioned to large scale secondary data, and those realizations are at the primary data scale. In this stage, the local conditioning well data is not exactly honored because the estimates are from the large scale secondary data. The estimate of each realization now will be integrated with the well data using Bayesian updating equation. Figure 3 illustrates the main idea of the method.

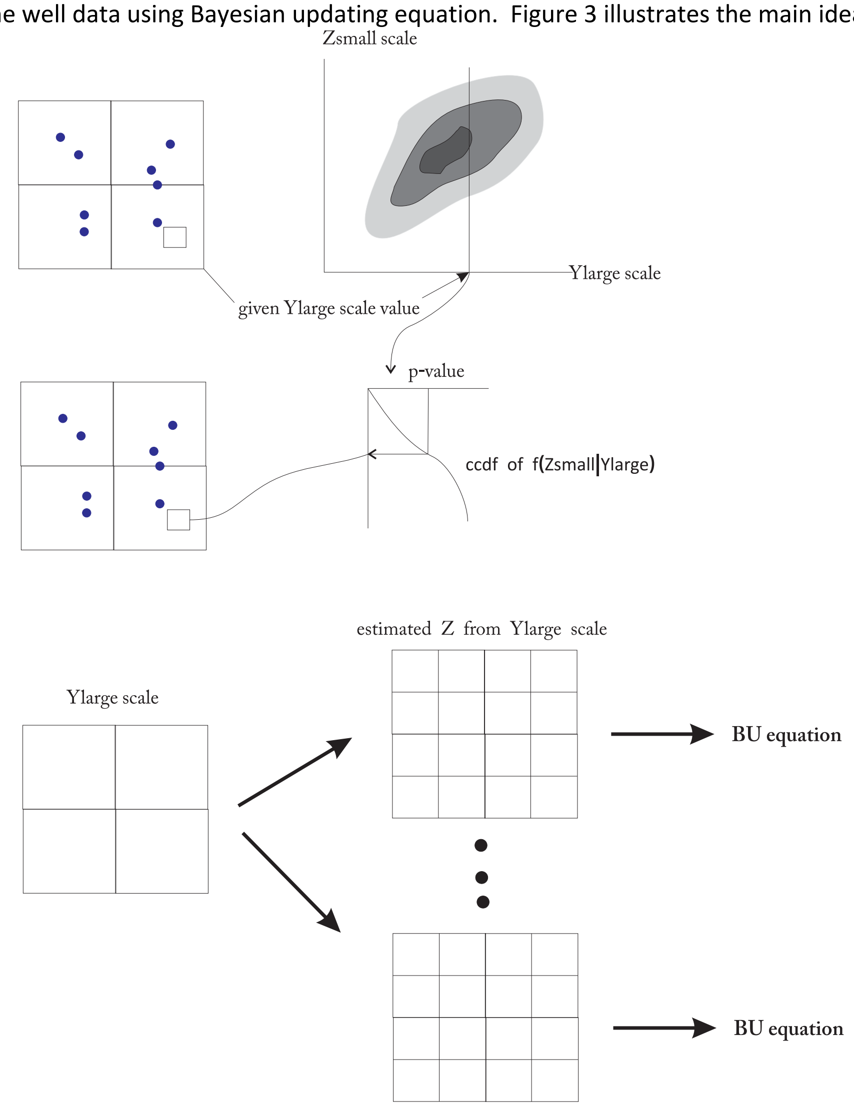


Figure 3: Proposed method of multiscale bivariate modeling applied to the Bayesian updating

3. Discussions

This note addresses the multiscale multivariate modeling. There are no known statistical references that mention the ineligibility of the multiscale multivariate modeling. Some of the previous reservoir modeling work implicitly assume the bivariate relation between two variables at different scales. This note shows potential applicability of multiscale multivariate modeling. The discussed method can be extended to the Bayesian updating for data integration. Although direct use of multiscale plotting and its summary statistics seem to be reasonable, care should be taken if there is large scale difference among the variables.

References

Efendiev, Y., Datta-Gupta, A., Ma, X., and Mallick, B., 2008, Modified Markov Chain Monte Carlo method for dynamic data integration using streamline approach, *Mathematical Geosciences*, 40: 213-232.

Lee, S., Malallah, A., Datta-Gupta, A., and Higdon, D., 2002, Multiscale data integration using Markov random fields, *Reservoir Evaluation and Engineering*,5:68-78.

Oz, B. and Deutsch, C.V., Size Scaling of Cross-Correlation between Multiple Variables, Natural Resources Research, 11(1), March 2002, pp 1-18.

Kupfersberger, H., Deutsch, C.V., and Journel, A.G. Deriving Constraints on Small-Scale Variograms due to Variograms of Large-Scale Data, Mathematical Geology, Vol. 30, No. 7, 1998, pp. 837-852.