COMMENT ON "ESTIMATING OR CHOOSING A GEOSTATISTICAL MODEL" BY OLIVER DUBRULE

MARK S. HANDCOCK
Department of Statistics & Operations Research
Leonard N. Stern School of Business
New York University
New York, NY 10012-1126 U. S. A.

Oliver Dubrule has presented an insightful and articulate view of the uncertainty quantification issues facing geostatisticians. These comments are intended to provide additional discussion.

The key feature of Bayesian approaches is to choose to quantify all uncertainty via probability distributions. This not only includes the variation in the phenomena, but extends to the uncertainty in the model and other "states of nature". This allows the geostatistician to quantify uncertainty in a natural and powerful way. Often the uncertainty represented is not solely the variability in the process of geological deposition, but also that of the geostatistician's lack of knowledge of that process. For example, the porosity of the rock at an given location is usually regarded as fixed and the model expresses the uncertainty in the geostatistician's knowledge of that value. Quantifying geological knowledge can be tricky and explicitly demands the involvement of the geostatistician and the geologist. Often this is because we are required to provide knowledge about aspects of the model or the phenomena that are not easy to express.

We do not get something for nothing, as Dubrule's sand connectivity example shows. Expressing lack of knowledge about the direction of the channels by a uniform distribution over the possible directions leads to simulated data with a unreasonably highly level of connectivity. The key point here is that we are able to recognize that the implications of this particular quantification of knowledge are not consistent with our geological knowledge: our intuition about the resultant map-view is much better than our intuition about the distributions of channel directions. We can exploit this to ensure our quantification of geological knowledge are much more consistent with our actual knowledge. Let X be the map-view of the sand connectivity and θ the direction of the channels. Let $f(X \mid \theta)$ be the distribution of map-views given (a distribution of) directions θ. In the example θ is taken to be uniform and the resulting map-views $(f(X \mid \theta))$ were overly connected relative to map-views consistent with our geological knowledge. However, note that

$$P(\ X \mid \pi) \ = \ \int_{\Theta} f(X \mid \theta)\pi(\theta)d\theta$$

where $P(\ X \mid \pi)$ is the (marginal) distribution of the map-view given the chosen

R. Dimitrakopoulos (ed.), Geostatistics for the Next Century, 15–16.

distribution $\pi(\theta)$ of θ. This expression makes it clear how the distribution of directions effects the resulting map-views. We can use it to play "what if" scenarios with particular distributions for the direction by choosing $\pi(\theta)$ and seeing if the resulting map-view is consistent with our knowledge. For example we could choose a distribution with subsequent directions highly correlated rather than independent. This would lead to less connected map-views. The distribution could be fine-tuned by comparing its implications in terms of map-views with our intuition about how the map-views should look.

In this example we see that "lack of knowledge" can be a slippery concept. Apparent ignorance from one perspective appears to be knowledge from another. The issue is how to elicit the knowledge. By "lack of knowledge" we often mean that it is difficult to specify any knowledge beyond that in the data. The objective then is to choose a model for model uncertainty that has minimal impact on the subsequent inference relative to the information in the data. This is often difficult, as **any** quantification of geological knowledge will say something precise about the phenomena (as this example shows). The solution is to check that the conclusions from the analysis are robust to the choice of quantification of knowledge. If they are not, then we have learned that the present data can not answer the question and we can not reach an unambiguous conclusion unless we augment the data by more information in the form of additional data or geological knowledge.