Spatial Prediction (Kriging) Revisited

- locally optimal (in least squares sense) prediction of unknown attribute values
- attribute data are reproduced, and prediction error statistics are reported
- spatially varying (data configuration dependent) smoothing

Problems:
- locally accurate predictions do not preserve: (i) histogram of original data, and (ii) their spatial structure (variogram model)
- smoothing effect could be detrimental if Kriging predictions are used as parameters in non-linear multiple-point models, e.g., visualization or flow simulation
- Kriging variance is not a complete measure of spatial attribute uncertainty

Remedy via stochastic simulation: (the other route being getting more or related data)
- impose reproduction of: (i) sample data, (ii) their histogram, and (iii) their spatial correlation (variogram)
- use in a Monte Carlo framework for uncertainty propagation/analysis

Example: Objective and Data

- calculate length $c$ of submarine cable required to connect locations A and B
- $y = [y(t_m), m = 1, \ldots, M]^T$: vector with $M = 200$ true sea floor depth values at equally spaced ($d = 1$ unit) intervals between A and B
- $y_s = [y(s_n), n = 1, \ldots, N]^T$: vector with $N = 15$ sample depth soundings

$$c = \sum_{m=1}^{M-1} \sqrt{[y(t_m) - y(t_{m+1})]^2 + d^2} = \text{662} \quad \text{depth at A and B}$$

True cable length $c$ = non-linear function of unknown depth profile $y$, i.e., $c = \phi(y)$
Example: Kriging-Derived Depth

Kriging of sea floor depth: \( \hat{y} = [\hat{y}(t_m), m = 1, \ldots, M]^T \)

- Simple Kriging using \( N = 15 \) sample data and true depth covariogram model:
  \[ \sigma_Y(h) = 0.05\delta_{h,0} + 0.95\exp(-3h^2/20^2) \]
- Kriging-derived sea floor depth reproduces (“passes through”) sample depth data
- predicted sea floor depth = smooth (generalized) version of true depth

\[
\hat{c} = \sum_{m=1}^{M-1} \sqrt{[\hat{y}(t_m) - \hat{y}(t_{m+1})]^2 + d^2} = 367
\]

Relative error of predicted cable length: \( \frac{\hat{c} - c}{c} = \frac{367 - 662}{662} \approx -45\% \ !!! \)

Example: One Conditional Depth Realization

Simulated sea floor depth: \( y_1 = [y_1(t_m), m = 1, \ldots, M]^T \)

- using \( N = 15 \) sample data and true depth covariogram model:
  \[ \sigma_Y(h) = 0.05\delta_{h,0} + 0.95\exp(-3h^2/20^2) \]
- simulated sea floor depth realization reproduces sample depth data
- simulated sea floor depth = approximate version of true depth

\[
c_1 = \sum_{m=1}^{M-1} \sqrt{[y_1(t_m) - y_1(t_{m+1})]^2 + d^2} = 631
\]

Relative error of simulated cable length: \( \frac{c_1 - c}{c} = \frac{631 - 662}{662} \approx -5\% \ !!! \)
Example: Multiple Conditional Depth Realizations

- numerous \((L)\) alternative versions of reality: \(Y = [y_1 \cdots y_I \cdots y_L]\)
- all \(L\) realizations reproduce measured depths at their respective sampling locations, as well as (approximately) the data histogram and variogram model
- ensemble (set) of \(L\) simulated depth realizations provides model of uncertainty regarding true sea floor depth

![Graph showing two alternative sea floor depth realizations](image)

no single realization can be regarded as truth; some are more likely than others

Use all \(L\) simulated realizations to propagate uncertainty in sea-floor depth to uncertainty in cable length

Example: Uncertainty in Cable Length

1. generate \(L\) simulated sea floor depth realizations \([y_1 \cdots y_I \cdots y_L]\)
2. compute \(L\) simulated cable lengths \([c_1 \cdots c_I \cdots c_L]\)
3. construct distribution of \(L\) simulated lengths

![Histogram showing distribution of cable lengths](image)

Risk-conscious prediction of cable length: Using: (i) the distribution of simulated cable lengths – this is called uncertainty analysis, and (ii) an appropriate loss (or profit) function – this is called cost-benefit analysis, (iii) compute the optimal prediction \(\hat{c}_{opt}\) for the true cable length \(c\) – this is called decision-making under uncertainty

Note: Uncertainty is independent of reported predictions
Example: Ensemble Average

- mean of simulated sea floor depth values at any target location $t_m$:

$$
E\{Y(t_m)|y_s\} \approx \frac{1}{L} \sum_{l=1}^{L} y_l(t_m), \quad m = 1, \ldots, M
$$

- one summary of simulated realizations; **not** a valid realization

- for multi-Gaussian data, ensemble average $E\{Y(t_m)|y_s\}$ of $L$ realizations approaches Kriging prediction $\hat{y}(t_m)$ as $L$ increases

---

Example: Ensemble Variance

- variance of simulated sea floor depth values at any location $t_m$:

$$
V\{Y(t_m)|y_s\} \approx \frac{1}{L} \sum_{l=1}^{L} [y_l(t_m) - E\{Y(t_m)|y_s\}]^2, \quad m = 1, \ldots, M
$$

- a measure of **local** (per-point) uncertainty regarding simulated values

- for multi-Gaussian data, ensemble variance $V\{Y(t_m)|y_s\}$ of $L$ realizations approaches (homoscedastic) Kriging variance as $L$ increases
**Example: White Noise Depth Realizations**

All $L$ realizations reproduce (approximately) the histogram (and possibly the data values at their sampling locations), *not* the covariogram model.

Classical Monte Carlo simulation from depth histogram is not enough !!!

*It is the histogram of the depth gradient that is needed*

Monte Carlo simulation from the depth gradient histogram would suffice.

---

**Example: Unconditional Depth Realizations**

All $L$ realizations reproduce (approximately) a histogram and covariogram model, *not* the data values at their sampling locations (correct shape or wavelength but random phase).

Uncertainty in true cable length is independent of phase of sea-floor depth !!!

*You just need a histogram and variogram model (not the data locations)*

*It is the histogram of the depth gradient that is needed*
Joint Spatial Uncertainty for Continuous Attributes

**Definition:** Uncertainty regarding \( M' \leq M \) unknown attribute values at a set of \( M' \leq M \) locations, given the sample data \( y_s \); e.g., probability that all \( M' \) values at \( M' \) locations be **simultaneously** no greater than an arbitrary threshold \( a \)

**Inference problem:** Joint \( M' \)-variate conditional CDF \( \mathbb{P}\{Y(t_1) \leq a, \ldots, Y(t_{M'}) \leq a | y_s\} \) can be evaluated analytically only for few multivariate distributions

**Solution:** evaluate above CDF numerically:

1. generate \( L \) simulated attribute realizations \( [y_l, l = 1, \ldots, L] \) at \( M' \) locations, conditional on the source data \( y_s \)
2. for each realization \( l \), compute an indicator \( i_l(a) = 1 \) if all \( M' \) attribute values are no greater than \( a \), \( i_l(a) = 0 \) if not.
3. Compute joint probability that all \( M' \) attribute values be simultaneously no greater than threshold \( a \) as the proportion (mean) of above indicators \( \frac{1}{L} \sum_{l=1}^{L} i_l(a) \)

Joint Spatial Uncertainty for Categorical Attributes

**Definition:** Uncertainty regarding \( M' \leq M \) unknown category codes at a set of \( M' \leq M \) locations, given sample codes; e.g., probability that all \( M' \) pixels be **simultaneously** black; \( M' \) black pixels constitute a pattern

**Key concepts:**
- set of \( M' \) single-point probabilities of code occurrence **not** sufficient for modeling joint spatial uncertainty, i.e., probability of pattern occurrence
- uncertainty in many spatially explicit model predictions = function of pattern occurrence, **not** of single-point (pixel-wise) code occurrences