# 地球統計学

## -Kriging and Simulation-

斎藤 広隆

Environmental Water Resources Engineering
Department of Civil & Environmental Engineering
The University of Michigan

# GEOSTATISTICS?

Georges Matheron (1962)

“ GEOSTATITICS, in their most general acceptation, are concerned with the study of **the distribution in space** of useful values for mining engineers and geologists, such as grade, thickness, or accumulation, including a most important practical application to the problems arising in ore-deposit evaluation”

**空間･時間に分布･変動するデータ(情報)を統計論的･決定論的に分析･モデリングするための道具**

# どのような分野で応用されているのか？

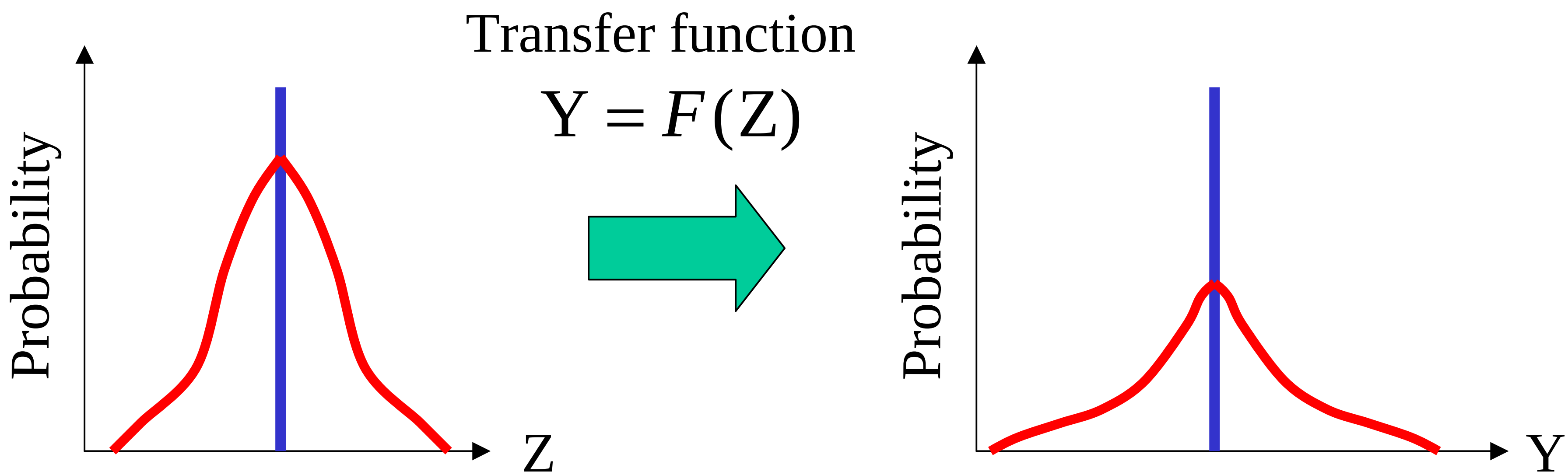
- 鉱物探査,石油工学,地球物理,地球化学
- 土壌科学,林学,農学
- 水文学,海洋学,気象学
- リモートセンシング,GIS,環境学
- 疫学

日本語訳：地球統計学 (新井, 1985)

# どんなジャーナルを見ればいいのか？

- Mathematical Geology, Computers & Geosciences
- Geoderma
- Soil Science, Soil Science Society of America Journal
- J. of Hydrology, WRR
- geoENV, GEOSATISTICS (proceedings)
- **本はたくさん出ている (注：日本語の本はなし)**

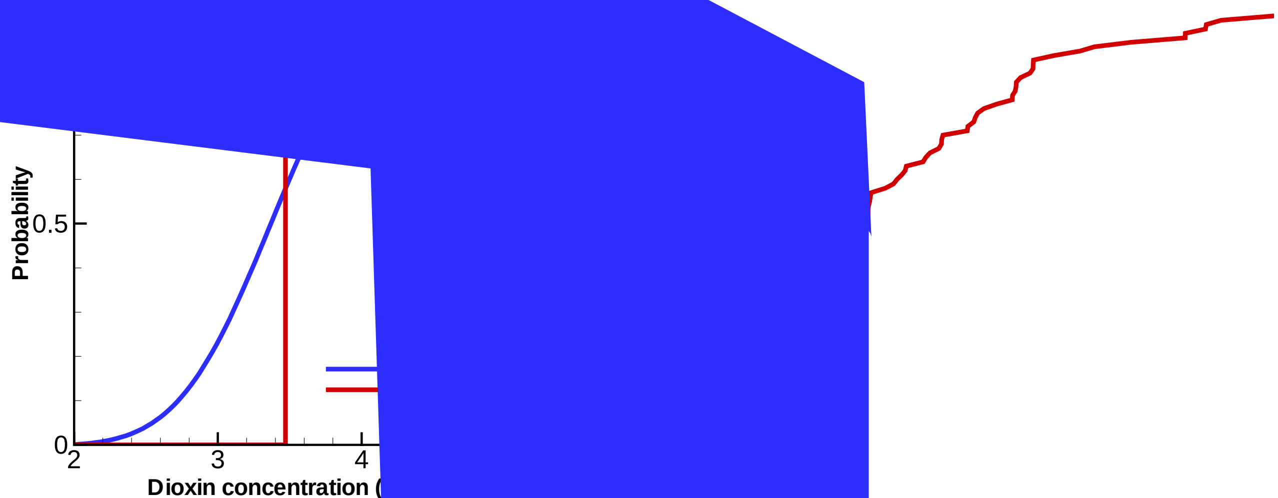
# Deterministic vs. Stochastic




**例：放射性廃棄物が貯蔵庫から浸透して地下水に到達する時間の推定**

# ulative Distribution Function

$$\Pr\{Z(\mathbf{u}) \leq z \mid (n)\}$$

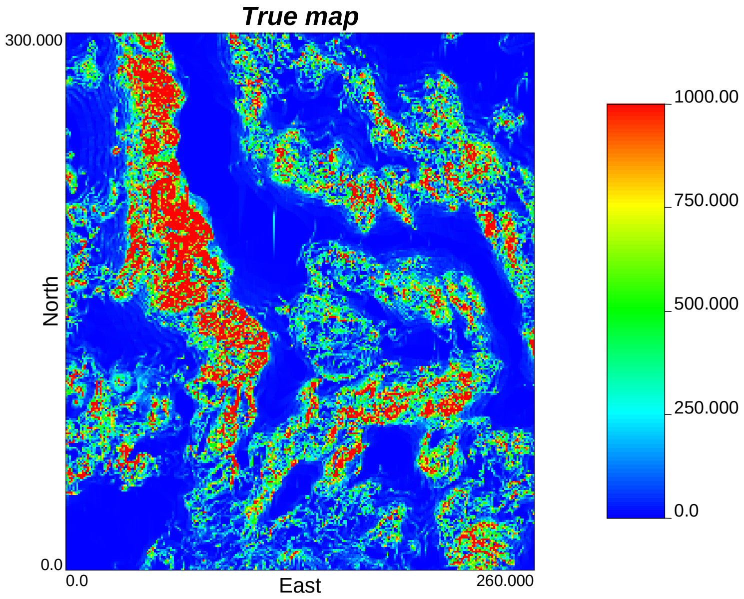


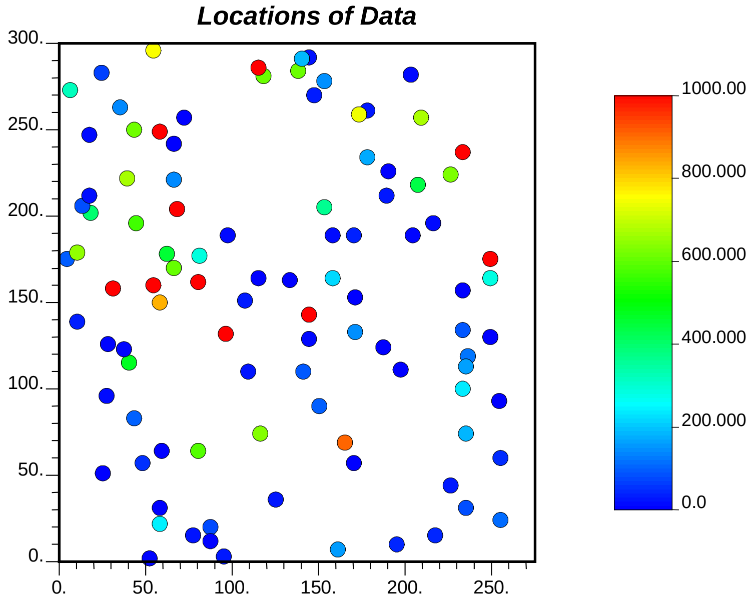
# GSLIB

## Geostatistical Software Library

- Programs developed at Stanford
  - SCRF: Stanford Center of Reservoir Forecasting (Andre G. Journel)
- All source codes available in ANSI Standard Fortran 77 (www.gslib.com)
- Public domain, open source, no support
- Manual with CD-Rom (*GSLIB User's Guide*)

# Walker Lake Data Set



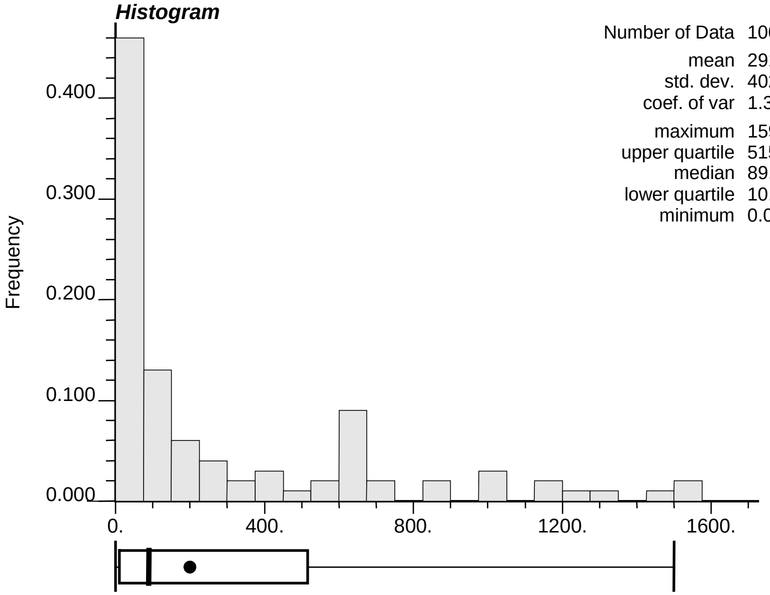




78000 points on 260 x 300 grid

100 points are randomly sampled

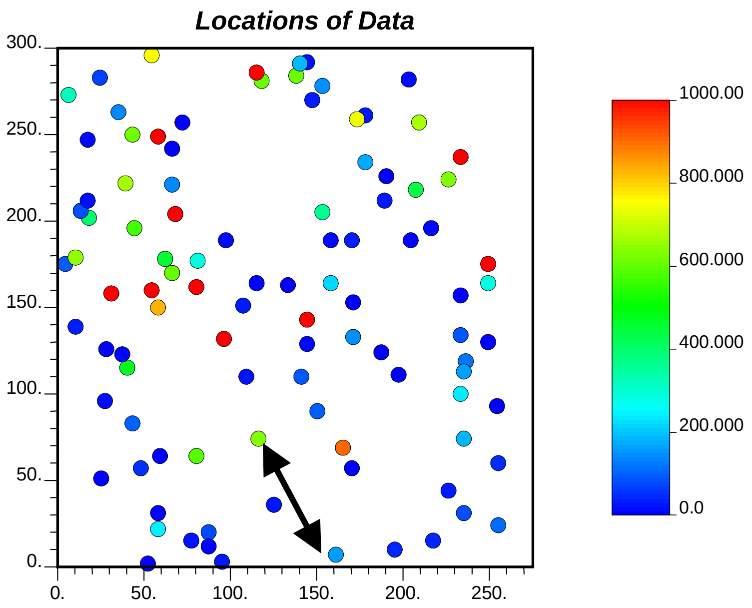
See Isaaks and Srivastava (1989) for details

# Univariate Description




- Summary statistics
- Highly skewed (median << mean)
- No spatial information!
- How to account for data locations?

# 空間分布の記述




- Location map**も空間分布の記述方法の一つ**
- **もっと定量的に表わす方法はないのか？**
- **ある点から離れると値はどれくらい変るのか？**

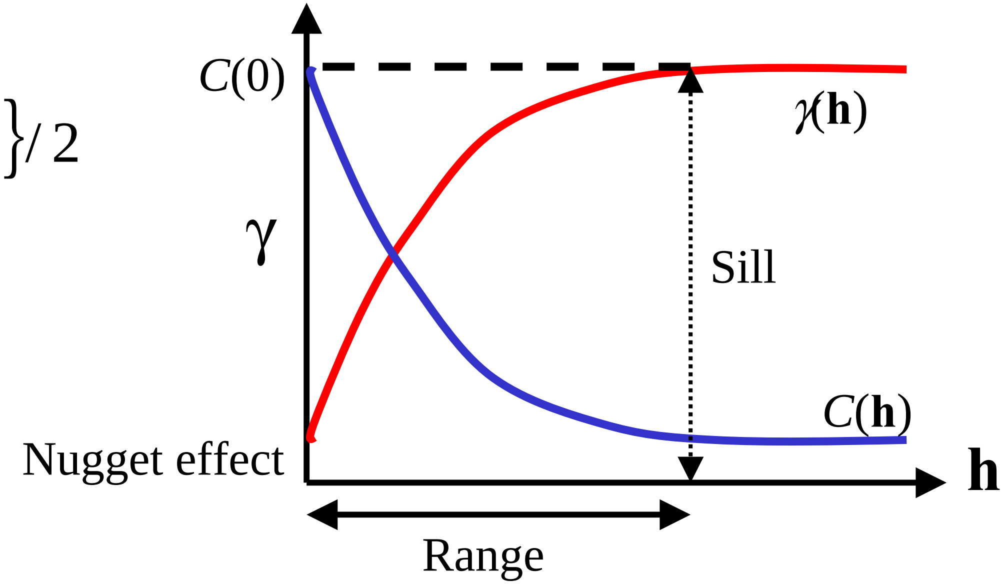
→ **データの相関を距離の関数として表わそう**

# Semivariogram

Measure of dissimilarity

$$\gamma(\mathbf{h}) = E\left\{[Z(\mathbf{u}) - Z(\mathbf{u}+\mathbf{h})]^2\right\}/2$$

**距離（h）のみの関数**




Correlogram

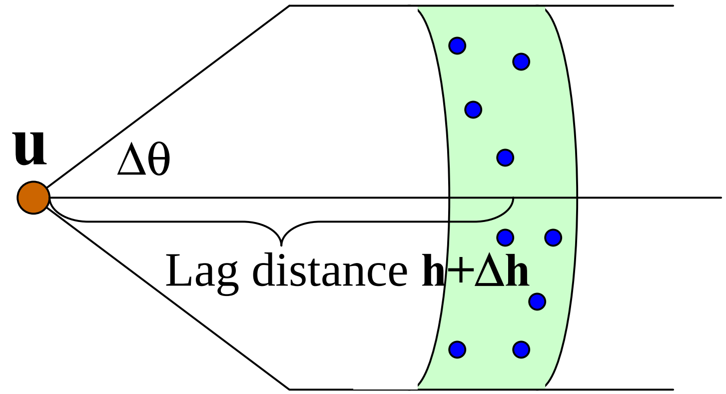
$$C(\mathbf{h}) = E\left\{Z(\mathbf{u}) \cdot Z(\mathbf{u}+\mathbf{h})\right\} - m^2$$

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$$

# Experimental Semivariogram

1. Compute experimental semivariogram (**hは実際はある区間を取る**)

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_\alpha) - z(\mathbf{u}_\alpha + \mathbf{h})]^2$$

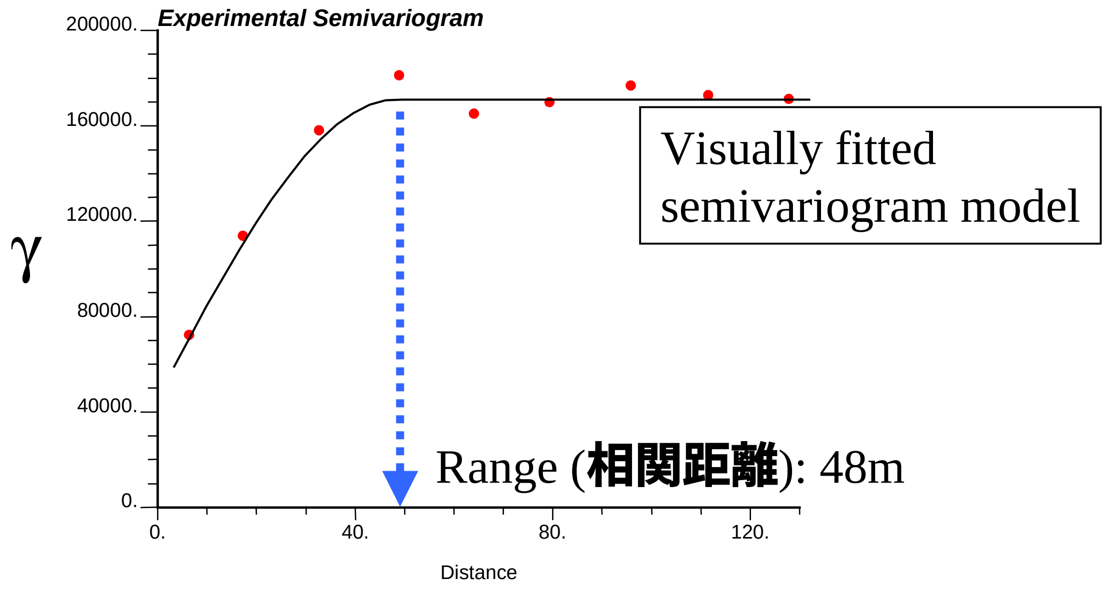



- **異方性の存在**

2. Fit permissible semivariogram models

- Automatic fitting
- Semi-automatic fitting

# Univariate Spatial Description

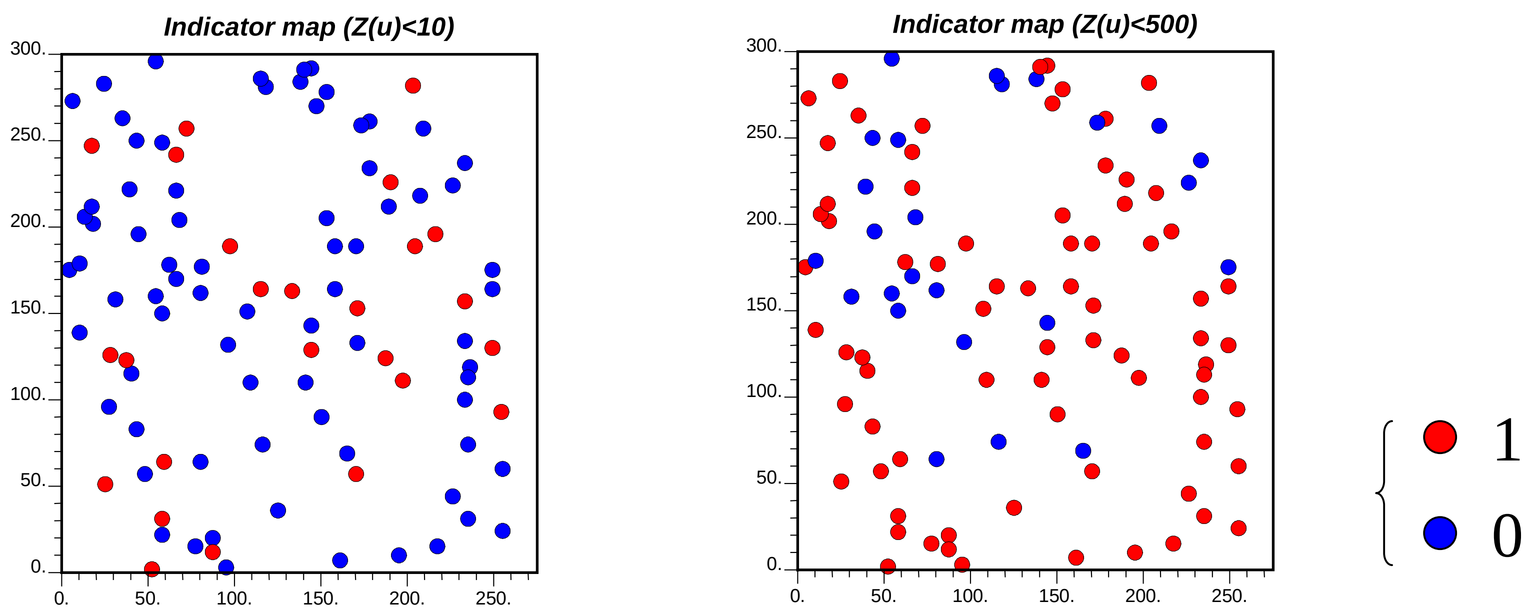



$$\gamma(\mathbf{h}) = 46{,}000 + 125{,}000\,\mathrm{Sph}\left(\frac{\mathbf{h}}{48}\right)$$

# Indicator (指示) Approach

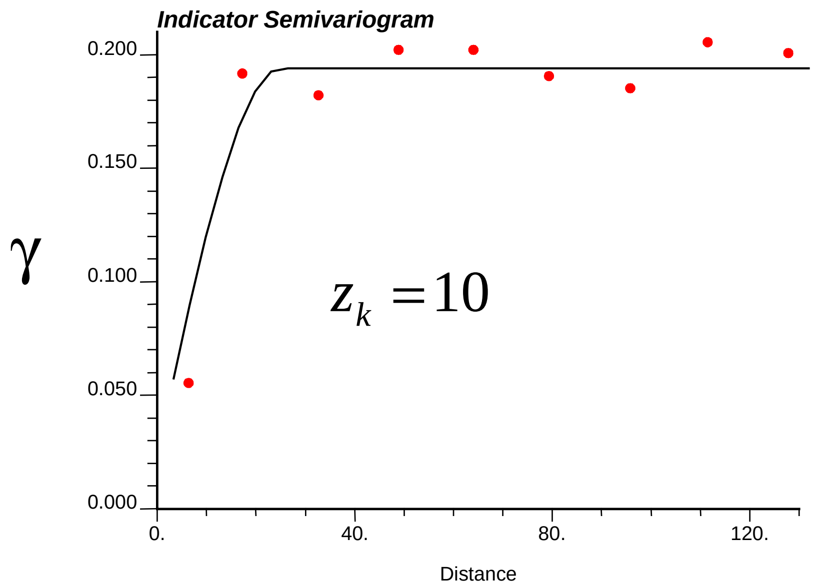
Spatial continuity of large or small values: Indicator approach

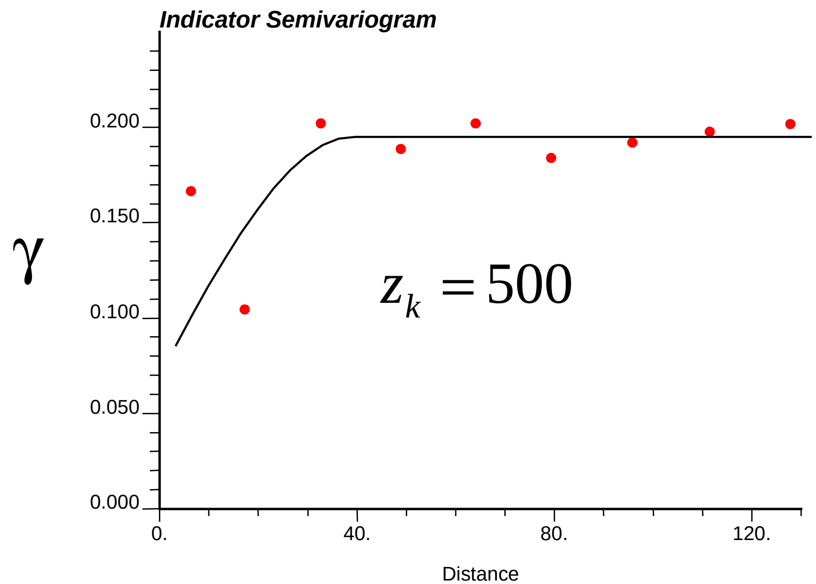
$$i(u_\alpha; z_k) = \begin{cases} 1 & \text{if } z(u_\alpha) \le z_k \\ 0 & \text{otherwise} \end{cases}$$



# Indicator Semivariogram

$$\gamma_I(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [i(\mathbf{u}_\alpha) - i(\mathbf{u}_\alpha + \mathbf{h})]^2$$



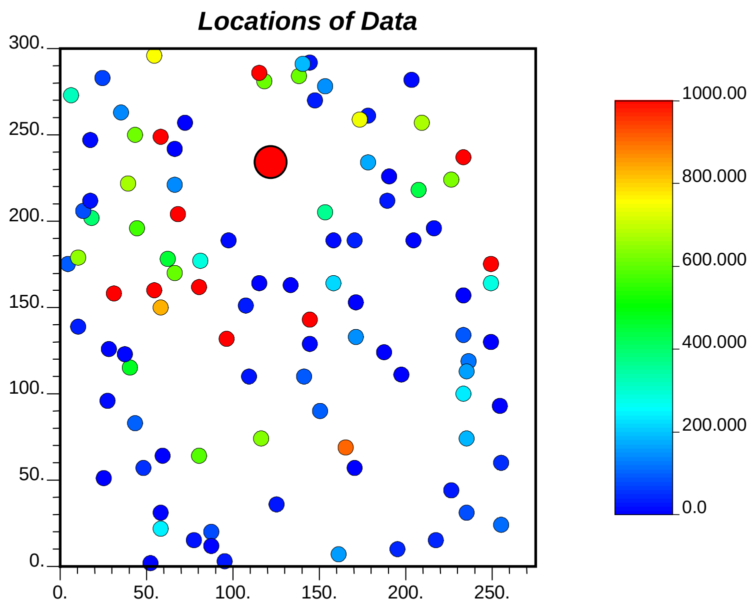

$$\gamma_I(\mathbf{h}) = 0.024 + 0.17\,\mathrm{Sph}\left(\frac{\mathbf{h}}{25}\right)$$

$$\gamma_I(\mathbf{h}) = 0.07 + 0.125\,\mathrm{Sph}\left(\frac{\mathbf{h}}{39}\right)$$

# 値の推定

## Estimation of the unknown value

Q. **以下の赤丸の場所の値は？**



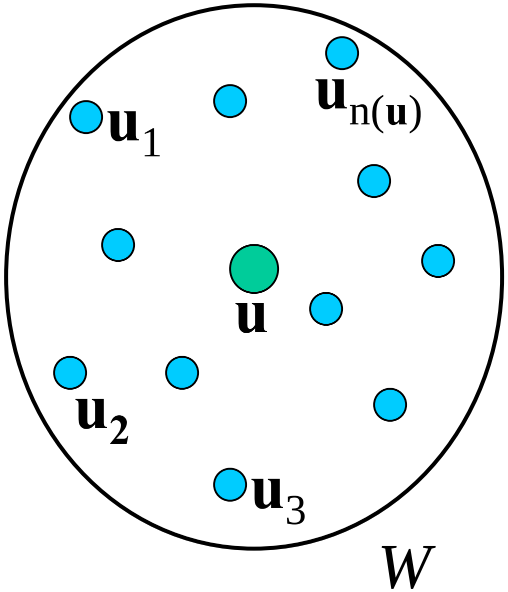

- Average + std. deviation
- Thiessen polygons
  - Estimated value = the closest observation
- Inverse distance
  - Linear combination of neighboring data
- Splines
  - Set of polynomials

No account of the data support, spatial variability, or estimation error!!

# Kriging (クリッギング)
## -Linear Regression Estimator-

$$\underbrace{z^*(\mathbf{u})}_{\text{Kriging estimate}} - \overbrace{m(\mathbf{u})}^{\text{Trend}} = \sum_{\alpha=1}^{n(\mathbf{u})} \overbrace{\lambda_\alpha(\mathbf{u})}^{\text{Kriging weight}} \left[ z(\mathbf{u}_\alpha) - m(\mathbf{u}_\alpha) \right]$$

- unknown $z(\mathbf{u})$, data $z(\mathbf{u}_\alpha)$
- $m(\mathbf{u}) = E[Z(\mathbf{u})]$
- $n(\mathbf{u})$ data within neighborhood $W(\mathbf{u})$



# Kriging Weights

**Objective**

Determine kriging weights $\lambda_\alpha$ such that estimation variance is minimized under the unbiasedness constraint

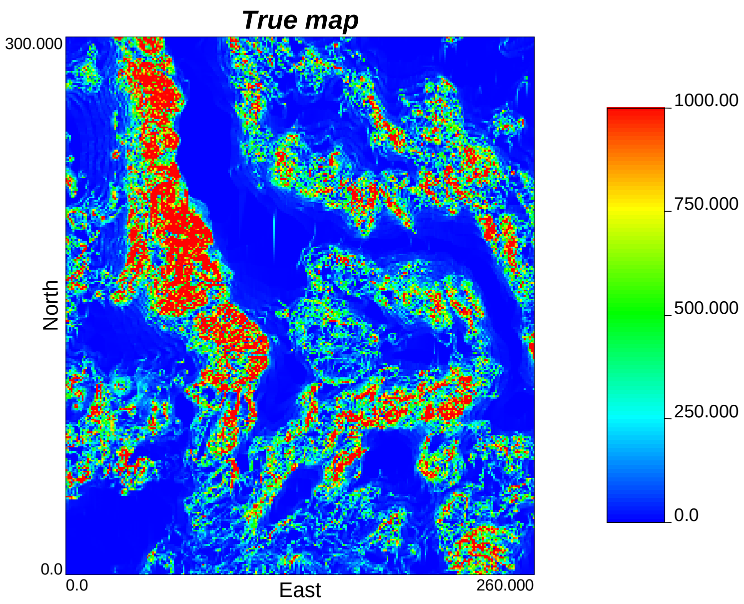
$$\begin{cases} \sigma_K^2(\mathbf{u}) = \mathrm{Var}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} \\ \mathrm{E}\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0 \end{cases}$$

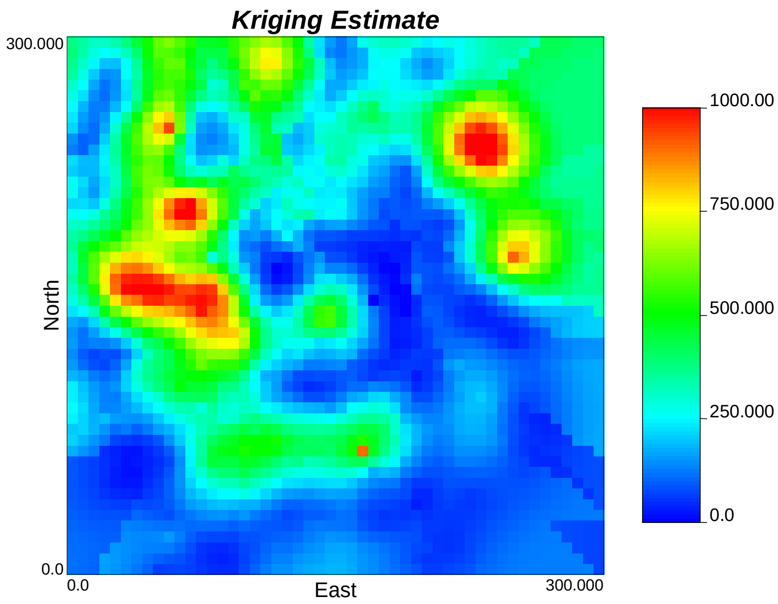
e.g. Kriging System (n($\mathbf{u}$) + 1 linear equations)

$$\begin{cases} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u})\gamma(\mathbf{u}_\alpha - \mathbf{u}_\beta) - \mu_{OK}(\mathbf{u}) = \gamma(\mathbf{u}_\alpha - \mathbf{u}) \quad \alpha = 1,\ldots,n(\mathbf{u}) \\ \sum_{\beta=1}^{n(\mathbf{u})} \lambda_\beta^{OK}(\mathbf{u}) = 1 \end{cases}$$

$$\mathbf{K}_{OK}\boldsymbol{\lambda}_{OK}(\mathbf{u}) = \mathbf{k}_{OK}$$

# Kriging Estimate



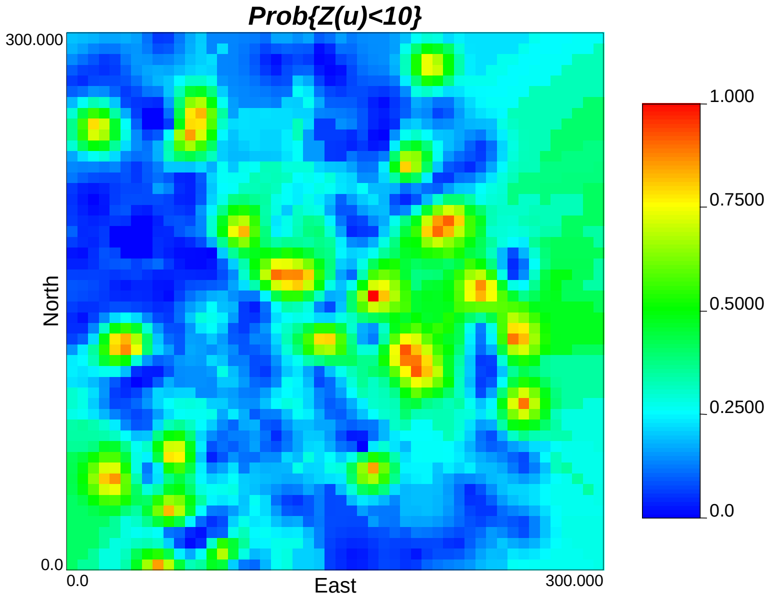

Estimated on 50 x 50 grid

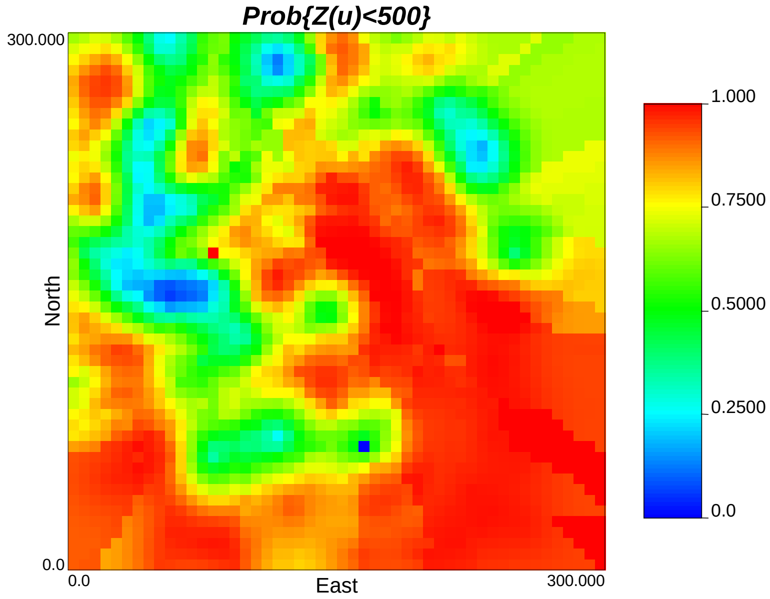
**いくつも違う種類のクリッギング法が開発されている.**

# Indicator Kriging

$$\Pr\{Z(u) \le z_k \mid (n)\} = \mathrm{E}\{I(u; z_k) \mid (n)\}$$

$$I(u_\alpha; z_k) = \begin{cases} 1 & \text{if } z(u_\alpha) \le z_k \\ 0 & \text{otherwise} \end{cases}$$




Probability of not exceeding a given threshold value.

# Kriging interpolation

- Smooth out details of the spatial variation.
- Small values are overestimated, while large values are underestimated.

Cannot assess spatial uncertainty.

# Dealing with Uncertainty

- 不均一なものをすべて必要なスケールで描写するのは不可能(もし可能であればDeterministicなモデルで十分)
- Stochasitc(確率論的) モデル
  - Monte Carlo Simulation：各確率変数にある確率密度(分布)関数を当てはめ、乱数を使いその分布から値をサンプリングする。1回のシミュレーションで得られる結果を“realization”と呼ぶ。シミュレーション毎に結果は異なっているので、100回、1000回とシミュレーションを走らせてその結果を統計的に処理する(例：平均でどれくらいの割合で10年以内に放射性物質が地下水に到達したか)。

# Geostatistical Stochastic Simulation

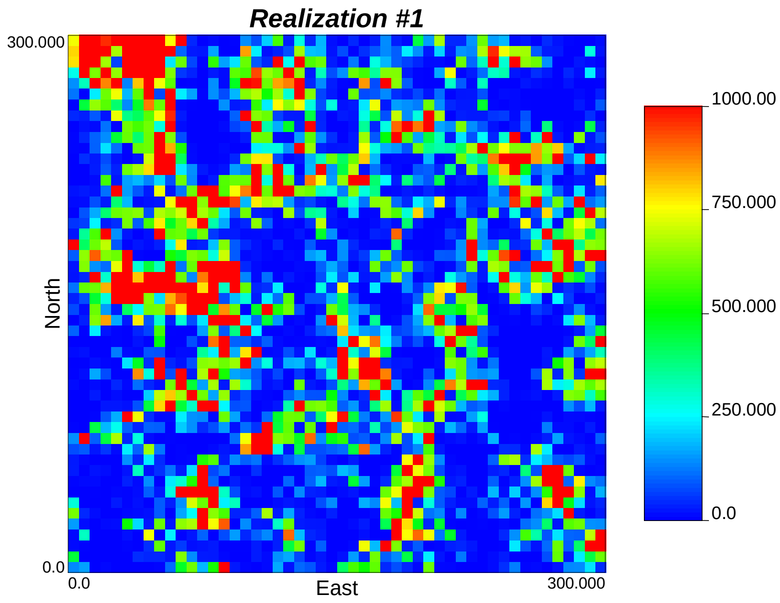
## -Monte Carlo Method-

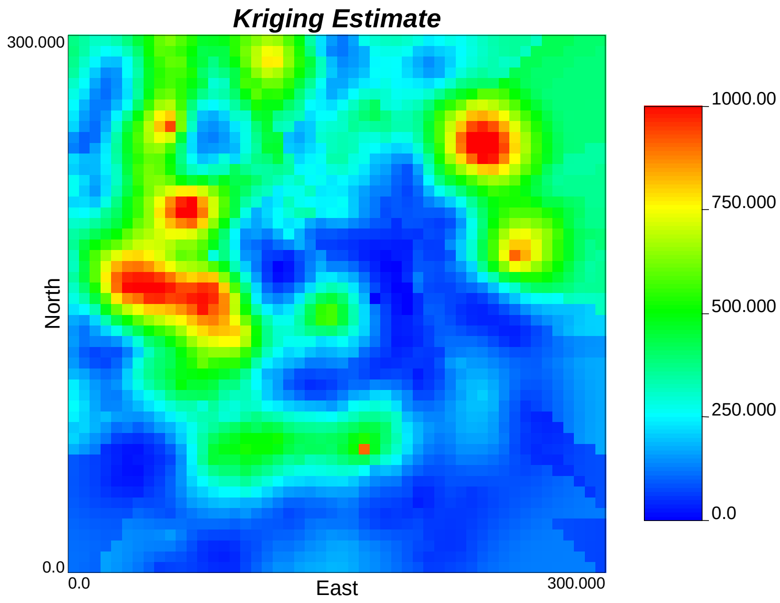
- Generate alternative representations (realizations or maps) of the spatial distribution of data values over the study area
- Conditional vs. unconditional

- Each realization (map) reproduces
  - Sample histogram (i.e. experimental variance)
  - Semivariogram model

**シミュレーションもクリッギングのように数多くの方法が開発されている.**

# Simulation vs. Estimation

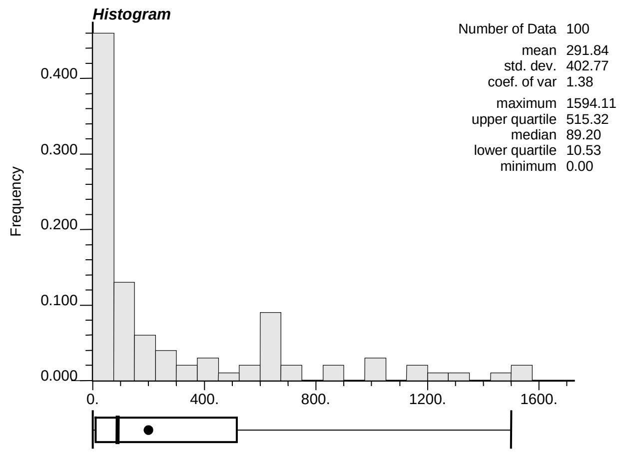


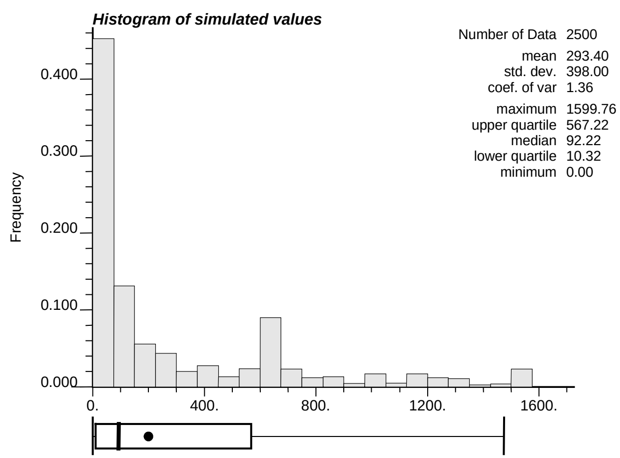



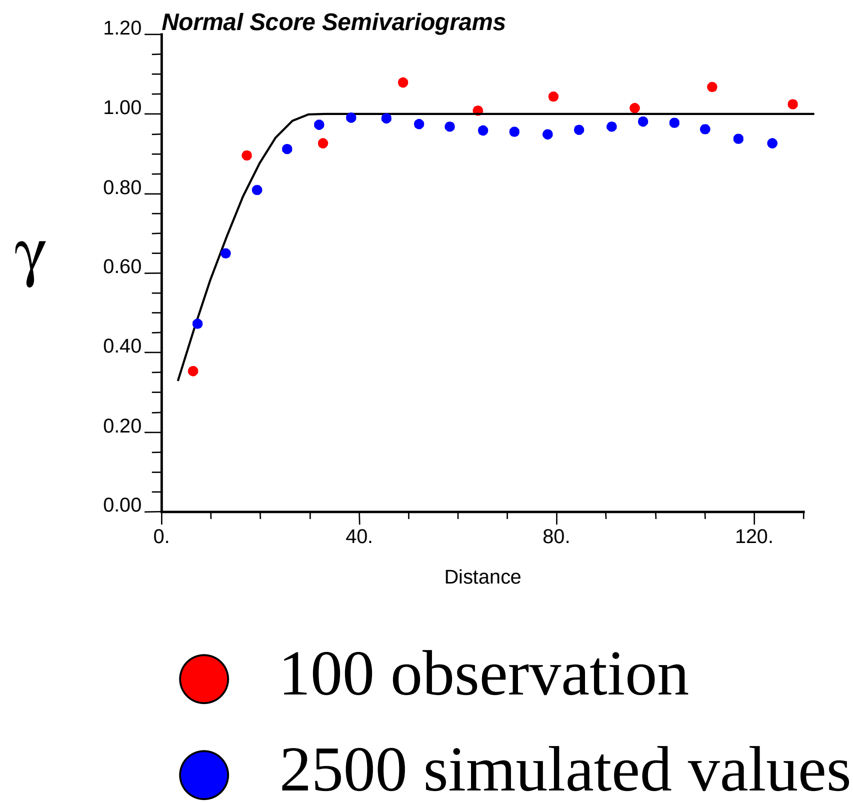

Simulated on 50 x 50 grid

Data values are honored at their locations (conditional simulation).

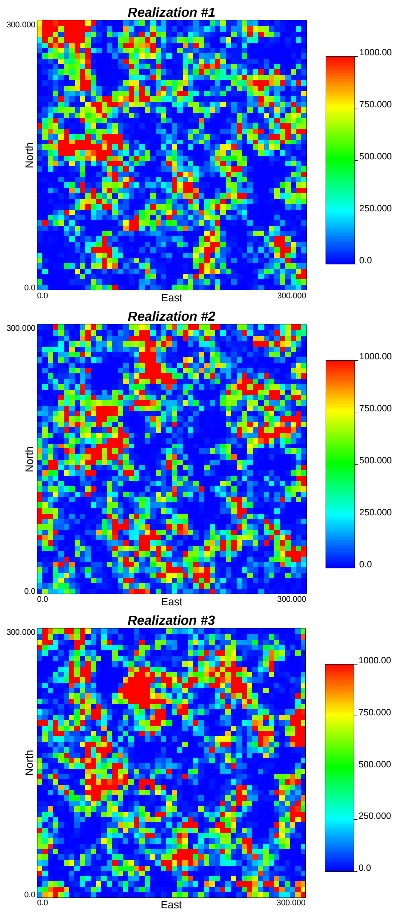
# Reproduction of Target Statistics



# Modeling the Spatial Uncertainty




Transfer function

$$Y = F(Z)$$

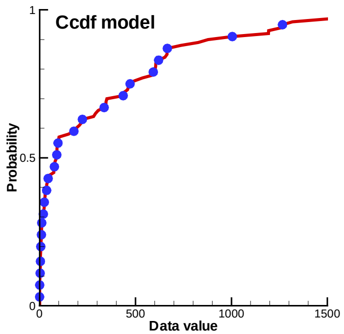
Each realization produces different output.

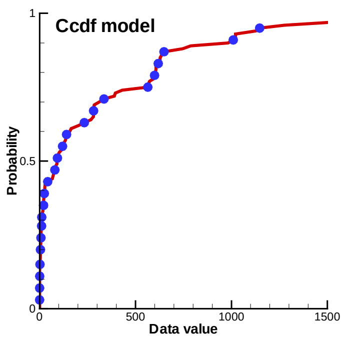
**それぞれのマップでの統計量はサンプルの統計量に等しい**

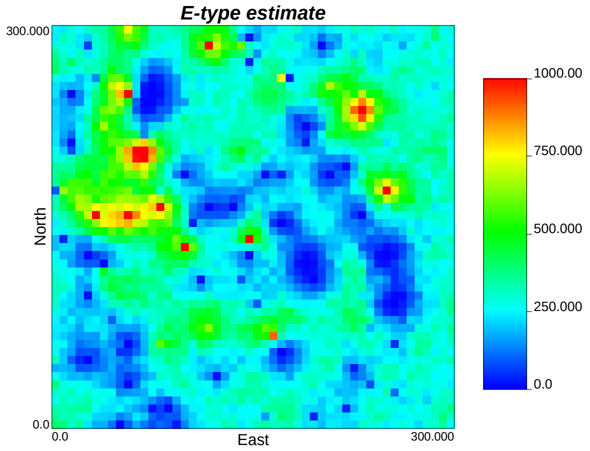
# Summary of Spatial Uncertainty

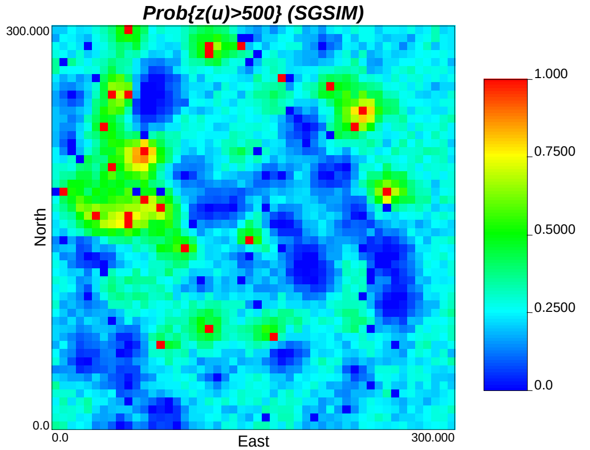
## Summary of 100 realizations of the spatial distribution of data

### Local uncertainty: ccdf



# シミュレーションの利点

- Propagation of uncertainty
  - Local/global transfer function
  - Quantify the uncertainty from input parameters to the model output
- Change of support
  - **測定したスケールからの**up scaling, down scaling

# Incorporation of Secondary Information: Multivariate Geostatistics

- **主変数以外の**2**次情報**(Secondary Information)**を使って推定やシミュレーションを行う**
  - **高価な測定＋安価な情報：透水係数の測定は金がかかるが、リモートセンシングデータは安くに手に入る。**
  - **測定誤差、汚染源の位置、事前情報などなど**

# 参考文献


- Deutsch, C. V., and A. G., Journel, *GSLIB: Geostatistical Software Library and Users Guide*, Oxford University Press, 1998
- Goovaerts, P., *Geostatistics for Natural Resources Evaluation*, Oxford University Press, 1997
- Goovaerts, P., “Geostatistical modeling of uncertainty in soil science”, *Gendarme*, Vol. 103, **2001**, pp 3-26
- Isaaks, E. H., and Srivastava, R. M., *An Introduction to Applied Geostatistics*, Oxford University Press, 1989
- Oliver, M. A., and R. Webster, “How geostatistics can help you”, *Soil Use and Management*, Vol. 7, No. 4, **1991**, pp 206-217
- Wackernagel, H., *Multivariate Geostatistics*, Springer, 1998


Questions?