Value of Information Analysis in Spatial Models

Jo Eidsvik

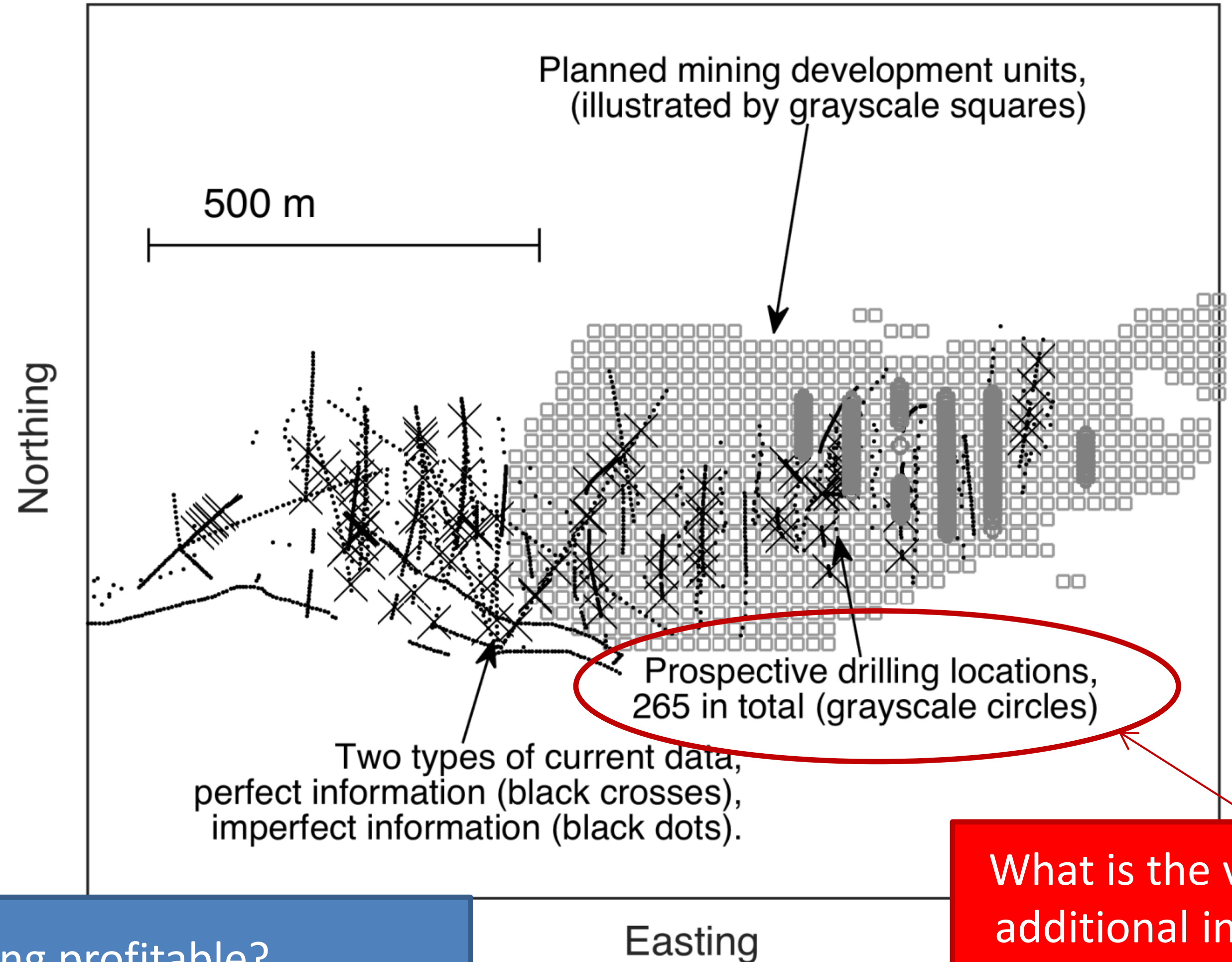
Jo.eidsvik@ntnu.no

Plan for course

Time	Topic
Monday	Introduction and motivating examples
	Elementary decision analysis and the value of information
Tuesday	Multivariate statistical modeling, dependence, graphs
	Value of information analysis for dependent models
Wednesday	Spatial statistics, spatial design of experiments
	Value of information analysis in spatial decision situations
Thursday	Examples of value of information analysis in Earth sciences
	Computational aspects
Friday	Sequential decisions and sequential information gathering
	Examples from mining and oceanography

Every day: Small exercise half-way, and computer project at the end.

Geostatistics (Kriging)



Is mining profitable?

What is the value of this additional information?

Joint pdfs

There are families of joint **pdfs**. Parametrically, or non-parametrically.

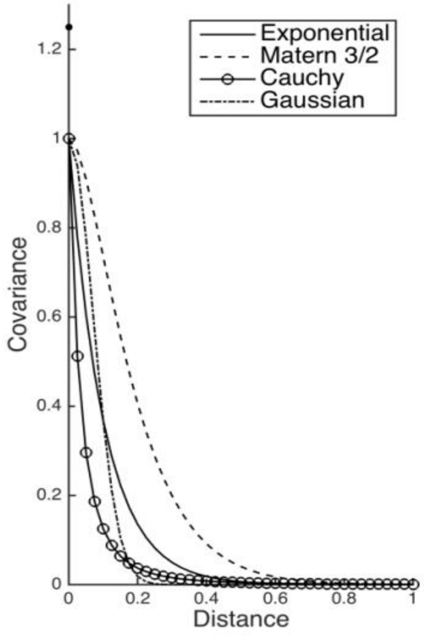
Gaussian distribution is very common:

$$p(\boldsymbol{x}) = N(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \ldots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & .. & \Sigma_{21} \\ .. & .. & .. & .. \\ \Sigma_{n1} & \Sigma_{n2} & .. & \Sigma_{nn} \end{pmatrix}$$

For a Gaussian process, in a spatial application, the covariance entries are formed in a particular way.

Spatial covariance functions

$$p(x) = N(\mu, \Sigma), \qquad \mathbf{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & .. & \Sigma_{21} \\ .. & .. & .. & .. \\ \Sigma_{n1} & \Sigma_{n2} & .. & \Sigma_{nn} \end{pmatrix} \qquad \Sigma_{ij} = \Sigma\left(\left|\mathbf{s}_i - \mathbf{s}_j\right|\right) = \Sigma\left(\left|\mathbf{t}\right|\right)$$



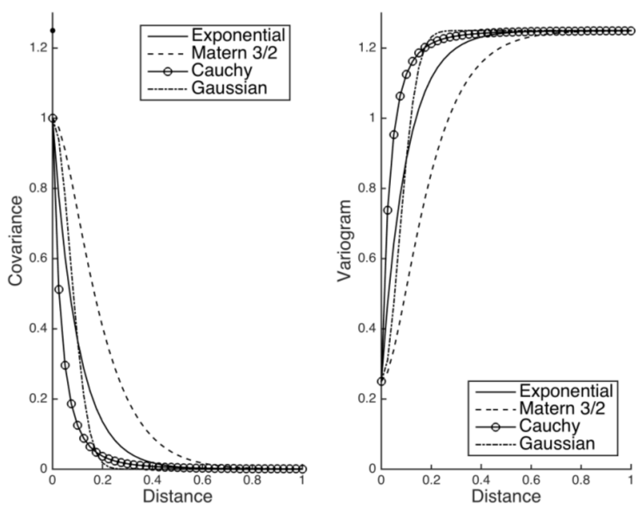
Model	Covariance
Exponential	$\Sigma(\lvert\mathbf{t}\rvert) = \sigma^2 \exp(-\eta \lvert\mathbf{t}\rvert)$
Matern 3/2	$\Sigma(\lvert\mathbf{t}\rvert) = \sigma^2 (1+\eta\lvert\mathbf{t}\rvert)\exp(-\eta\lvert\mathbf{t}\rvert)$
Cauchy-type	$\Sigma(\lvert\mathbf{t}\rvert) = \sigma^2 \frac{1}{(1+\eta\lvert\mathbf{t}\rvert)^3}$
Gaussian	$\Sigma(\lvert\mathbf{t}\rvert) = \sigma^2 \exp\left(-\eta^2 \lvert\mathbf{t}\rvert^2\right)$

Example - Gaussian process

$$p(\boldsymbol{x}) = N(\boldsymbol{\mu}, \Sigma), \qquad \boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + N(\boldsymbol{0}, \tau^2 \boldsymbol{I})$$

$$p(\boldsymbol{y}) = N(\boldsymbol{F}\boldsymbol{\mu}, \boldsymbol{C}) \qquad \boldsymbol{C} = \boldsymbol{F}\Sigma\boldsymbol{F}^t + \tau^2 \boldsymbol{I}$$

Design matrix: $\boldsymbol{F}$



Model	Covariance
Exponential	$C(\lvert\mathbf{t}\rvert) = \tau^2 I(\lvert\mathbf{t}\rvert = 0) + \sigma^2 \exp(-\eta\lvert\mathbf{t}\rvert)$
Matern 3/2	$C(\lvert\mathbf{t}\rvert) = \tau^2 I(\lvert\mathbf{t}\rvert = 0) + \sigma^2 (1+\eta\lvert\mathbf{t}\rvert)\exp(-\eta\lvert\mathbf{t}\rvert)$
Cauchy-type	$C(\lvert\mathbf{t}\rvert) = \tau^2 I(\lvert\mathbf{t}\rvert = 0) + \sigma^2 \frac{1}{(1+\eta\lvert\mathbf{t}\rvert)^3}$
Gaussian	$C(\lvert\mathbf{t}\rvert) = \tau^2 I(\lvert\mathbf{t}\rvert = 0) + \sigma^2 \exp(-\eta^2\lvert\mathbf{t}\rvert^2)$

Gaussian process - model

$$p(\boldsymbol{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \qquad \boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + N\left(\boldsymbol{0}, \tau^2 \boldsymbol{I}\right)$$

$$p(\boldsymbol{y}) = N(\boldsymbol{F}\boldsymbol{\mu}, \boldsymbol{C}) \qquad \boldsymbol{C} = \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t + \tau^2 \boldsymbol{I}$$

Goal is:

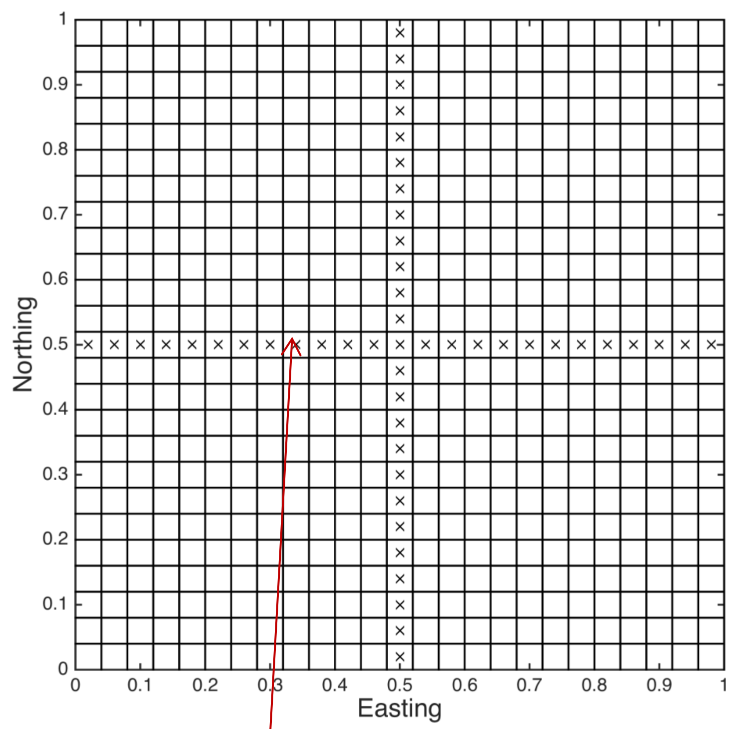
$$p(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{x})\, p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = N\left(\boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{F}^t\left(\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t + \tau^2\boldsymbol{I}\right)^{-1}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}), \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{F}^t\left(\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F}\boldsymbol{\Sigma}\right)$$

Norwegian wood - forestry example

Farmer must decide whether to harvest forest, or not. There is uncertainty about timber volumes and profits over the spatial domain.

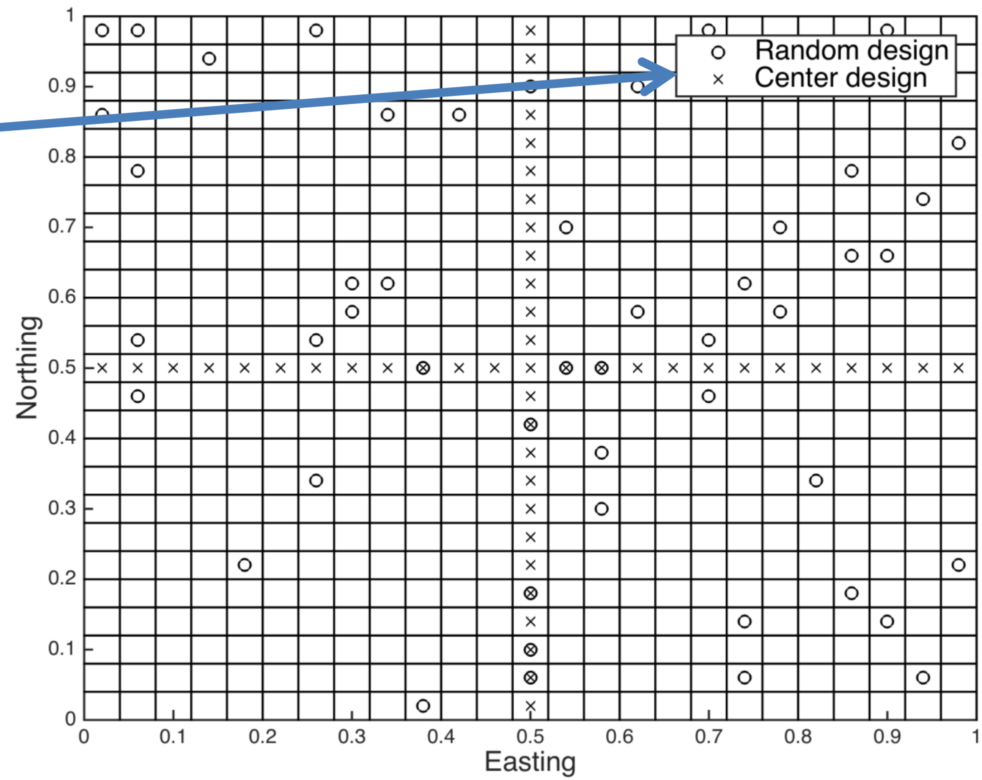
Another decision is whether to collect data before making these decisions.
If so, how and where should data be gathered.



Where to put survey lines for timber volumes information?
Typically partial, imperfect information.

Norwegian wood - posterior

Design matrix: $\boldsymbol{F}$



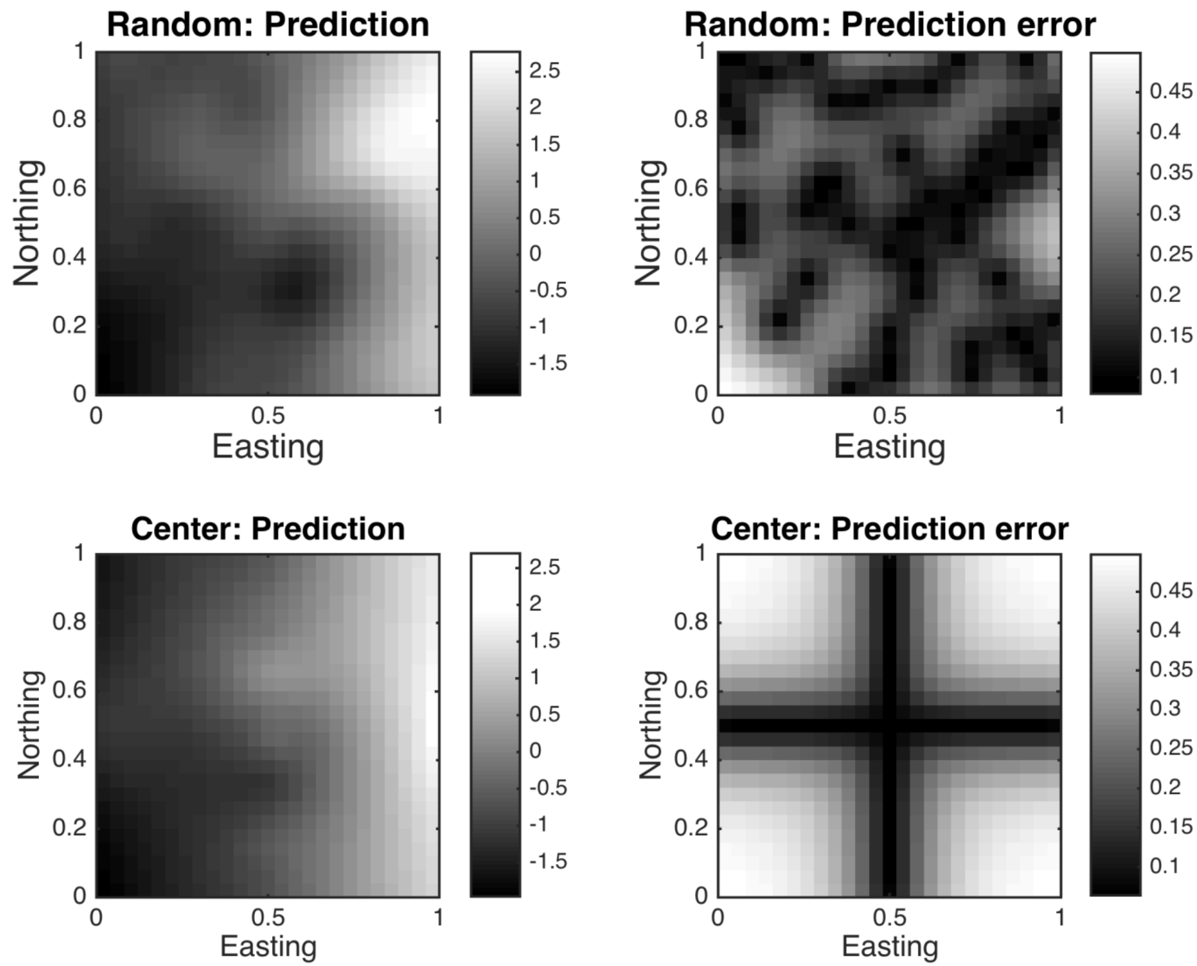
Goal is:

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{x})\, p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = N\left(\boldsymbol{\mu} + \boldsymbol{\Sigma F}^t \left(\boldsymbol{F\Sigma F}^t + \tau^2 \boldsymbol{I}\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}\right), \boldsymbol{\Sigma} - \boldsymbol{\Sigma F}^t \left(\boldsymbol{F\Sigma F}^t + \tau^2 \boldsymbol{I}\right)^{-1} \boldsymbol{F\Sigma}\right)$$

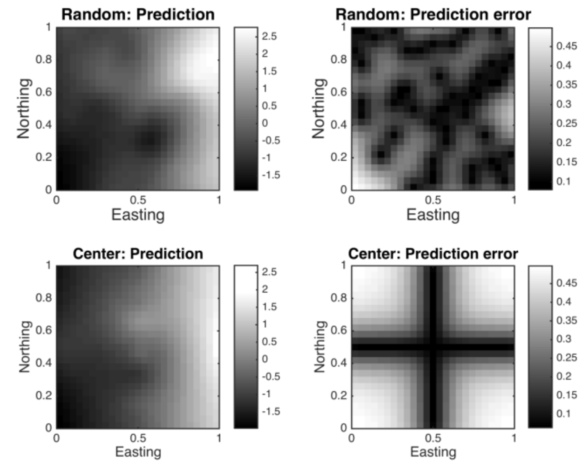
This is Kriging prediction and associated variance.

Norwegian wood – posterior results



$$p(x \mid y) = N\left(\boldsymbol{\mu} + \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2 \boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu}\right), \boldsymbol{\Sigma} - \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2 \boldsymbol{I}\right)^{-1}\boldsymbol{F\Sigma}\right)$$

Norwegian wood – information



We can base data gathering schemes on different criteria

- Maximum variance reduction
- Maximum entropy
- Value of information (VOI)

VOI is based on decision situation!
Others are not material – not tied to decision situation.

Spatial design

- Geometric criterion (space-filling design).
  - Minimize average distance between data locations.
  - Set a threshold on minimum distance to nearest data location.

  Challenging to compare various data accuracies.
- Variance reduction criterion.
- Kriging-related criteria (slope and weight of mean).
- Entropy reduction criterion.
- Prediction error.

Variance reduction

$$\mathrm{SV} = \sum_{i=1}^{n} Var(x_i) = \sum_{i=1}^{n} \Sigma_{ii} = \mathrm{trace}\left(\boldsymbol{\Sigma}\right)$$

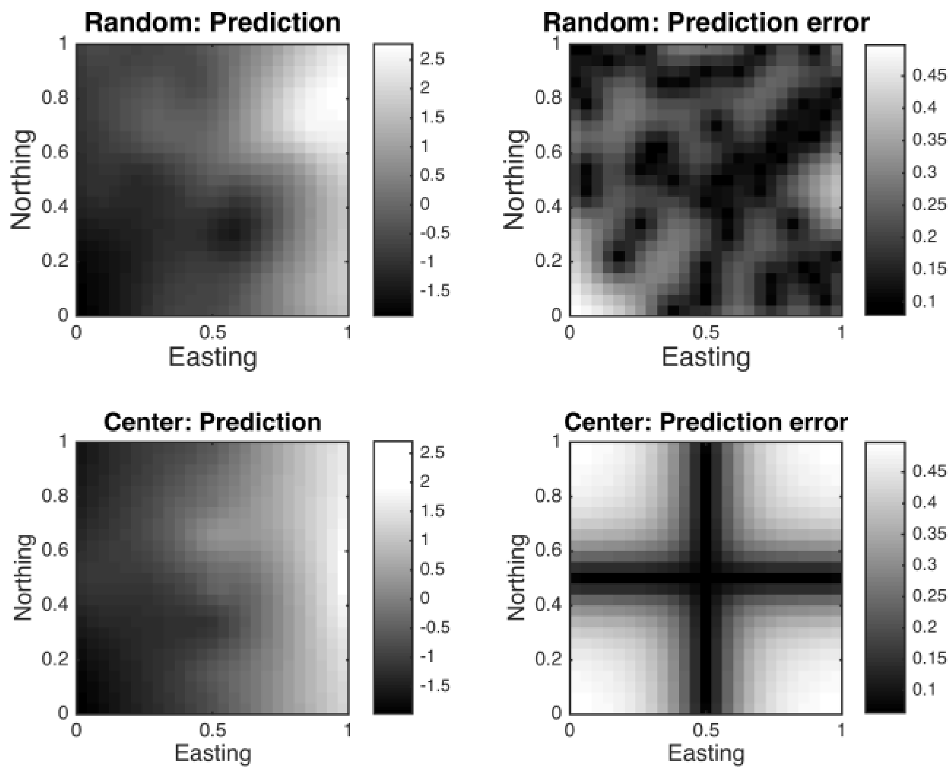
Expected variance reduction:

$$\mathrm{EVR}(\boldsymbol{y}) = \sum_{i=1}^{n} Var(x_i) - E\left(\sum_{i=1}^{n} Var(x_i \mid \boldsymbol{y})\right) = \sum_{i=1}^{n} Var(x_i) - \int \sum_{i=1}^{n} Var(x_i \mid \boldsymbol{y}) p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

Could use a weighted sum, or choose a subset of variables for prediction.

Variance reduction (Kriging)

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = N\left(\boldsymbol{\mu} + \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}(\boldsymbol{y} - \boldsymbol{F\mu}), \boldsymbol{\Sigma} - \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F\Sigma}\right)$$



Overall variance reduction is larger for the random design.

Other criteria tied to Kriging

$$p(\boldsymbol{x} \mid \boldsymbol{y}) = N\left(\boldsymbol{\mu} + \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\left(\boldsymbol{y} - \boldsymbol{F\mu}\right), \boldsymbol{\Sigma} - \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F\Sigma}\right)$$

Test function: $\boldsymbol{w}^t$

- **Slope of regression** : Regression between the predicted and true (block) variables.

$$Cov\left(\boldsymbol{w}^t\boldsymbol{x}, \boldsymbol{w}^t\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}\right) \Big/ Var\left(\boldsymbol{w}^t\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{y}}\right)$$

- **Weight of the mean** : The relative impact (for blocks), of regression versus that of Kriging.

$$\frac{\boldsymbol{w}^t\left(\boldsymbol{\mu} - \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F\mu}\right)\boldsymbol{1}}{\boldsymbol{w}^t\left(\boldsymbol{\mu} - \boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F\mu}\right)\boldsymbol{1} + \boldsymbol{w}^t\left(\boldsymbol{\Sigma F}^t\left(\boldsymbol{F\Sigma F}^t + \tau^2\boldsymbol{I}\right)^{-1}\right)\boldsymbol{1}}$$

Entropy (Shannon)

$$Ent(\boldsymbol{x}) = -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}$$

$$Ent(\boldsymbol{x} \mid \boldsymbol{y}) = -\int p(\boldsymbol{x} \mid \boldsymbol{y}) \log p(\boldsymbol{x} \mid \boldsymbol{y}) d\boldsymbol{x}$$

Expected mutual information:

$$\mathrm{EMI}\left(\boldsymbol{y}\right) = Ent(\boldsymbol{x}) - \int Ent(\boldsymbol{x} \mid \boldsymbol{y})\, p(\boldsymbol{y}) d\boldsymbol{y}$$

Entropy of a Gaussian

$$Ent(\boldsymbol{x}) = -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}$$

$$Ent\left(\boldsymbol{x}\right) = \frac{n}{2}\left(1 + \log\left(2\pi\right)\right) + \frac{1}{2}\log\left|\boldsymbol{\Sigma}\right|$$

Very commonly used in the design of spatial experiments
(air quality monitoring, river monitoring networks, etc.)

Entropy of a Gaussian

Exercise:

1. **Compute the entropy of a bivariate standard Gaussian (with no correlation).**
2. **Adjust the correlation so that a Gaussian bivariate model with variance equal to 1.5 gets the same entropy as the standard Gaussian in 1..**

$$\boldsymbol{\Sigma} = 1.5\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

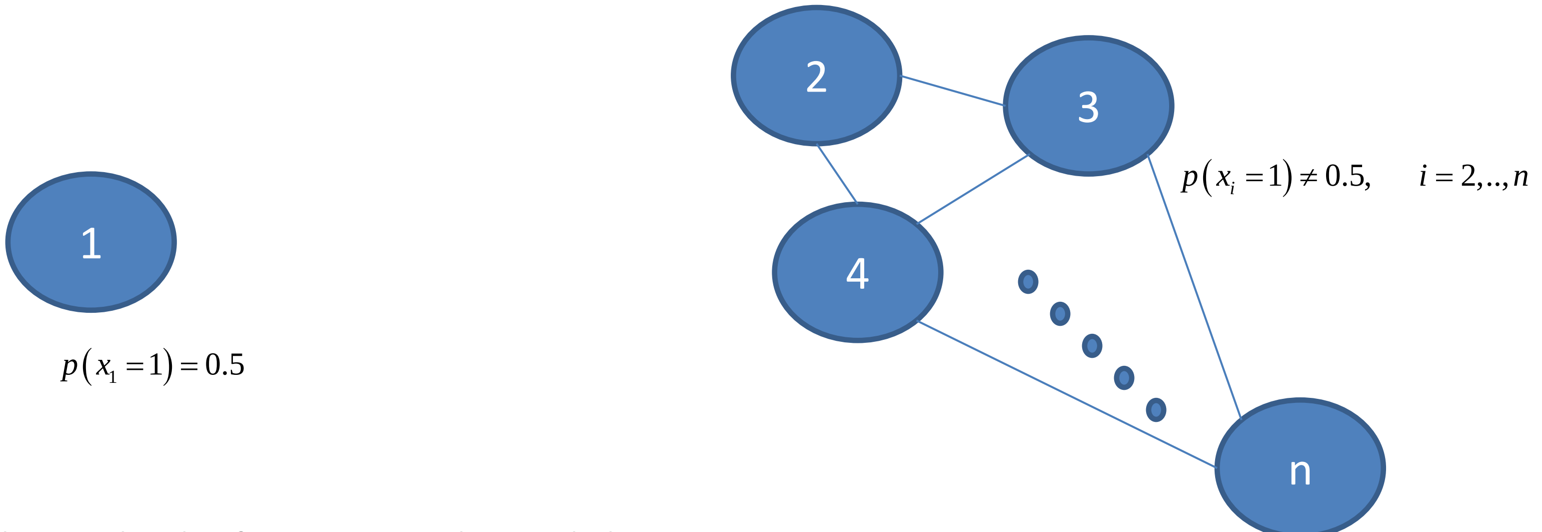
$$Ent(\boldsymbol{x}) = -\int p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}$$

$$Ent(\boldsymbol{x}) = \frac{n}{2}\left(1 + \log(2\pi)\right) + \frac{1}{2}\log|\boldsymbol{\Sigma}|$$

Entropy blind-spot

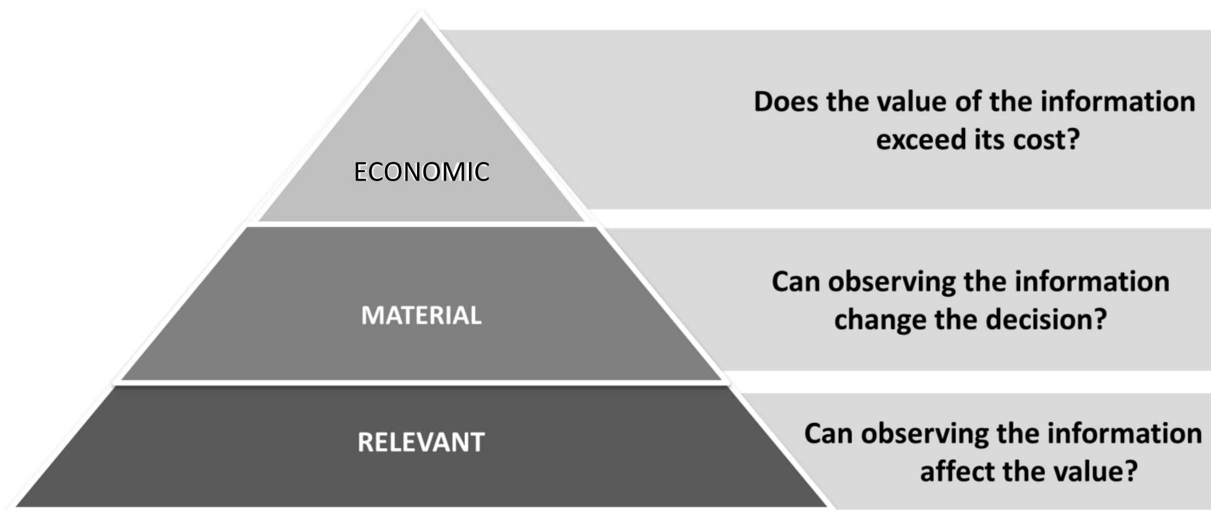
$$EMI(x_{\mathbb{K}}) = Ent(\boldsymbol{x}) - \int Ent(\boldsymbol{x}_{\mathbb{L}} \mid x_{\mathbb{K}})\, p(x_{\mathbb{K}})\, dx_{\mathbb{K}},$$

$$= Ent(x_{\mathbb{K}}) + \int Ent(\boldsymbol{x}_{\mathbb{L}} \mid x_{\mathbb{K}})\, p(x_{\mathbb{K}})\, dx_{\mathbb{K}} - \int Ent(\boldsymbol{x}_{\mathbb{L}} \mid x_{\mathbb{K}})\, p(x_{\mathbb{K}})\, dx_{\mathbb{K}} = Ent(x_{\mathbb{K}}).$$



This criterion always looks for marginals with largest entropy!
For a single observation, entropy would select node 1!

VOI - Pyramid of conditions



Pyramid of conditions - VOI is different from other information criteria (entropy, variance, prediction error, etc.)

Formula for VOI

$$PV = \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ E\left(v(\boldsymbol{x}, \boldsymbol{a}) \right) \right\} = \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ \int_{\boldsymbol{x}} v(\boldsymbol{x}, \boldsymbol{a}) p\left(\boldsymbol{x} \right) d\boldsymbol{x} \right\}$$

$$PoV\left(\boldsymbol{y} \right) = \int \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ E\left(v\left(\boldsymbol{x}, \boldsymbol{a} \right) \mid \boldsymbol{y} \right) \right\} p\left(\boldsymbol{y} \right) d\boldsymbol{y}$$

$$VOI\left(\boldsymbol{y} \right) = PoV\left(\boldsymbol{y} \right) - PV.$$

Formula for VOI

Spatial alternatives.

Spatial uncertainties.

$$PV = \max_{a \in A} \left\{ E\left(v(\mathbf{x}, \mathbf{a}) \right) \right\} = \max_{a \in A} \left\{ \int_{\mathbf{x}} v(\mathbf{x}, \mathbf{a}) \, p\left(\mathbf{x} \right) d\mathbf{x} \right\}$$

Spatial value function.

$$PoV\left(\mathbf{y} \right) = \int \max_{a \in A} \left\{ E\left(v\left(\mathbf{x}, \mathbf{a} \right) \mid \mathbf{y} \right) \right\} p\left(\mathbf{y} \right) d\mathbf{y}$$

Spatial data.

$$VOI\left(\mathbf{y} \right) = PoV\left(\mathbf{y} \right) - PV.$$

Information gathering

	Perfect	Imperfect
Total	Exact observations are gathered for all locations. This is rare, occurring when there is extensive coverage and highly accurate data gathering. $$\boldsymbol{y} = \boldsymbol{x}$$	Noisy observations are gathered for all locations. This is common in situations with remote sensors with extensive coverage, e.g. seismic, radar, satellite data. $$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}$$
Partial	Exact observations are gathered at some locations. This might occur, for instance, when there is careful analysis of rock samples along boreholes in a reservoir or a mine. $$\boldsymbol{y}_{\mathbb{K}} = \boldsymbol{x}_{\mathbb{K}}, \qquad \mathbb{K} \text{ subset}$$	Noisy observations are gathered at some locations. Examples include hand-held (noisy) meters to observe grades in mine boreholes, electromagnetic testing along a line, biological surveys of species, etc. $$\boldsymbol{y}_{\mathbb{K}} = \boldsymbol{x}_{\mathbb{K}} + \boldsymbol{\varepsilon}_{\mathbb{K}}, \qquad \mathbb{K} \text{ subset}$$

Decision situations and values

	Assumption: Decision Flexibility	**Assumption: Value Function**
Low decision flexibility; Decoupled value	Alternatives are easily enumerated $$a \in A$$	Total value is a sum of value at every unit $$v(\boldsymbol{x}, a) = \sum_j v(x_j, a)$$
High decision flexibility; Decoupled value	None $$\boldsymbol{a} \in \boldsymbol{A}$$	Total value is a sum of value at every unit $$v(\boldsymbol{x}, \boldsymbol{a}) = \sum_j v(x_j, a_j)$$
Low decision flexibility; Coupled value	Alternatives are easily enumerated $$a \in A$$	None $$v(\boldsymbol{x}, a)$$
High decision flexibility; Coupled value	None $$\boldsymbol{a} \in \boldsymbol{A}$$	None $$v(\boldsymbol{x}, \boldsymbol{a})$$

Decoupling – values are sums

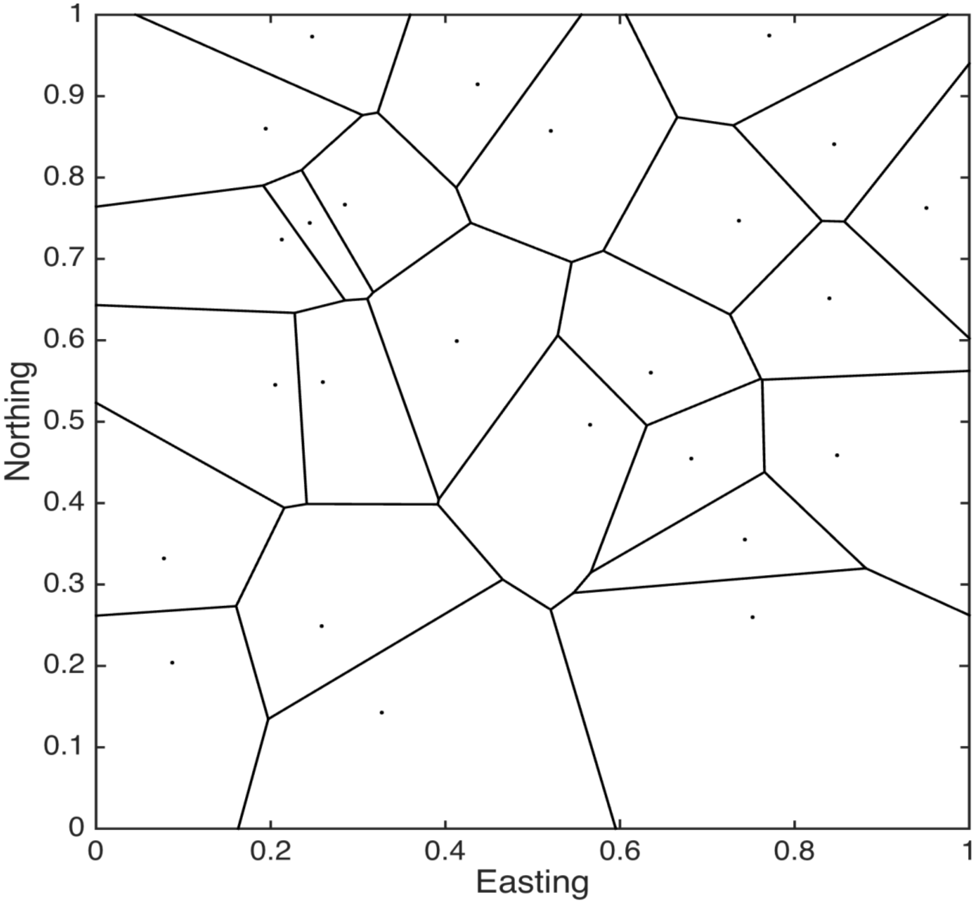
	Assumption: Decision Flexibility	**Assumption: Value Function**
Low decision flexibility; Decoupled value	Alternatives are easily enumerated $$a \in A$$	Total value is a sum of value at every unit $$v(\mathbf{x},a)=\sum_j v(x_j,a)$$
High decision flexibility; Decoupled value	None $$\boldsymbol{a} \in \boldsymbol{A}$$	Total value is a sum of value at every unit $$v(\mathbf{x},\boldsymbol{a})=\sum_j v(x_j,a_j)$$
Low decision flexibility; Coupled value	Alternatives are easily enumerated $$a \in A$$	None $$v(\mathbf{x},a)$$
High decision flexibility; Coupled value	None $$\boldsymbol{a} \in \boldsymbol{A}$$	None $$v(\mathbf{x},\boldsymbol{a})$$

Profit is sum of timber volumes from units.

Low versus high decision flexibility

High flexibility:
Farmer can select individual forest units.

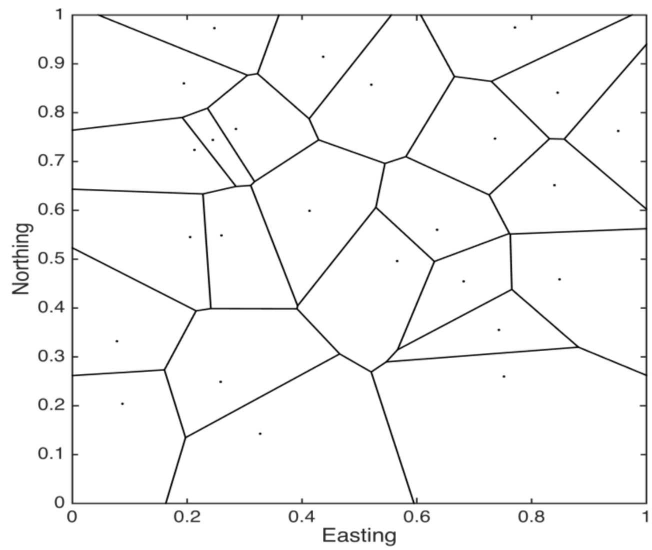
Low flexibility:
Farmer must select all forest units, or none.



Decoupled versus coupled value

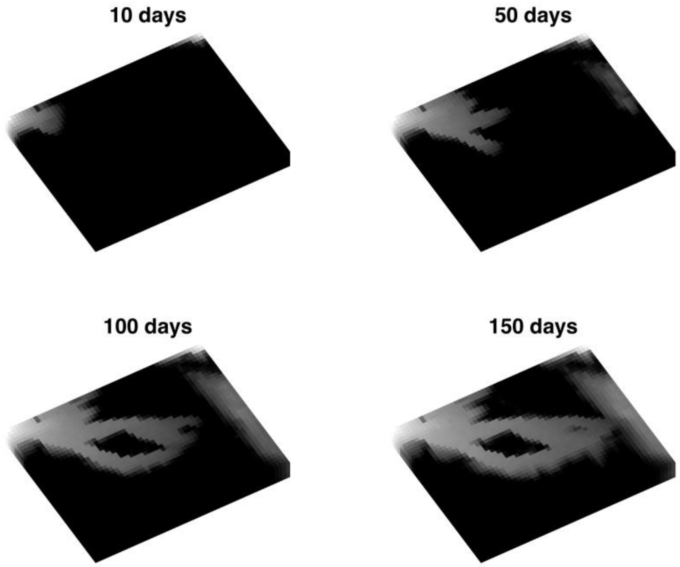
Farmer must decide whether to harvest at forest units, or not.

Value decouples to sum over units.



Petroleum company must decide how to produce a reservoir.

Value involves complex coupling of drilling strategies, and reservoir properties.



Computation - Formula for VOI

$$PV = \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ E\left(v(\boldsymbol{x}, \boldsymbol{a}) \right) \right\} = \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ \int_{\boldsymbol{x}} v(\boldsymbol{x}, \boldsymbol{a}) p\left(\boldsymbol{x} \right) d\boldsymbol{x} \right\}$$

$$PoV\left(\boldsymbol{y} \right) = \int \max_{\boldsymbol{a} \in \boldsymbol{A}} \left\{ E\left(v\left(\boldsymbol{x}, \boldsymbol{a} \right) | \boldsymbol{y} \right) \right\} p\left(\boldsymbol{y} \right) d\boldsymbol{y}$$

Computations :

- Easier with low decision flexibility (less alternatives).
- Easier if value decouples (sums or integrals split).
- Easier for perfect, total, information (upper bound on VOI).
- Sometimes analytical solutions. Otherwise approximations and Monte Carlo.

Formula for total perfect information

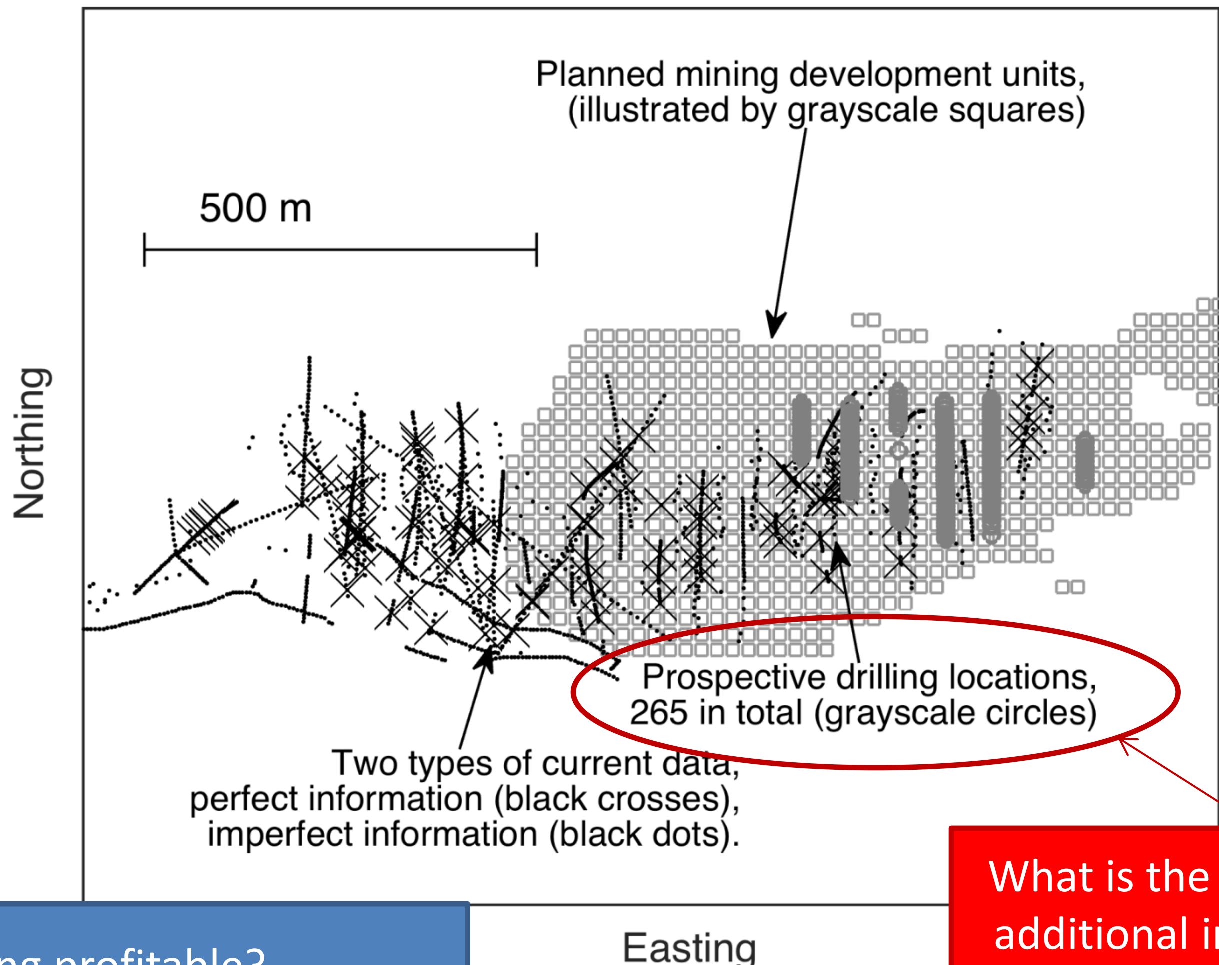
$$PV = \max_{\boldsymbol{a}\in \boldsymbol{A}} \left\{ E\left(v(\boldsymbol{x},\boldsymbol{a})\right)\right\} = \max_{\boldsymbol{a}\in \boldsymbol{A}} \left\{ \int_{\boldsymbol{x}} v(\boldsymbol{x},\boldsymbol{a})\, p\left(\boldsymbol{x}\right) d\boldsymbol{x} \right\}$$

$$PoV\left(\boldsymbol{x}\right) = \int \max_{\boldsymbol{a}\in \boldsymbol{A}} \left\{ v\left(\boldsymbol{x},\boldsymbol{a}\right)\right\} p\left(\boldsymbol{x}\right) d\boldsymbol{x}$$

$$VOI\left(\boldsymbol{x}\right) = PoV\left(\boldsymbol{x}\right) - PV.$$

Upper bound on any
information gathering scheme.

I love rock and ore – mining example

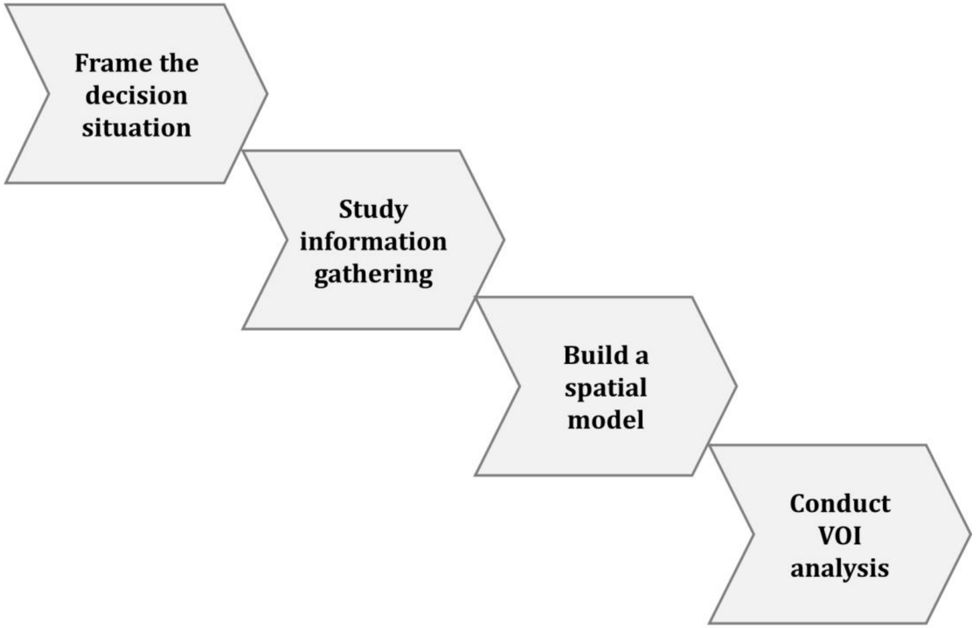


What is the value of this additional information?

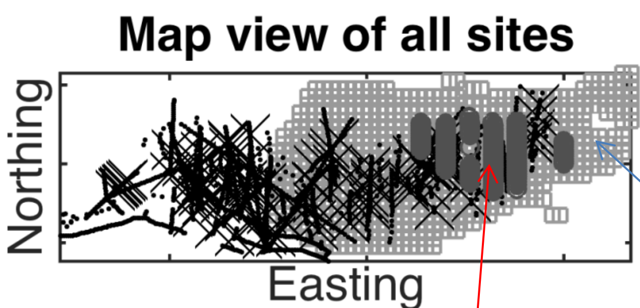
Is mining profitable?

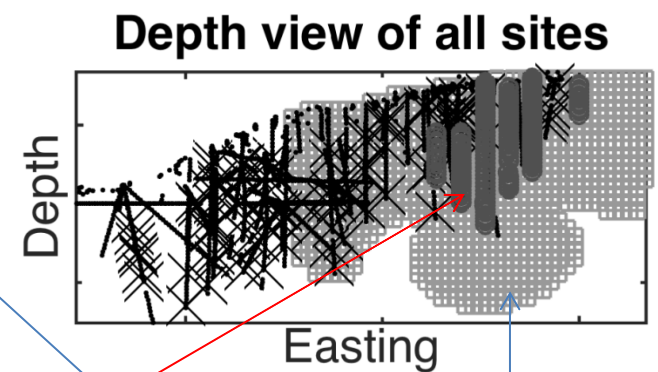
VOI workflow

- Low decision flexibility. De-coupled value function.
- Gather information by XRF or XMET in boreholes. No opportunities for adaptive testing.
- Model is a spatial Gaussian process.
- VOI analysis done by exact, Gaussian, computations.



Decision situation and data

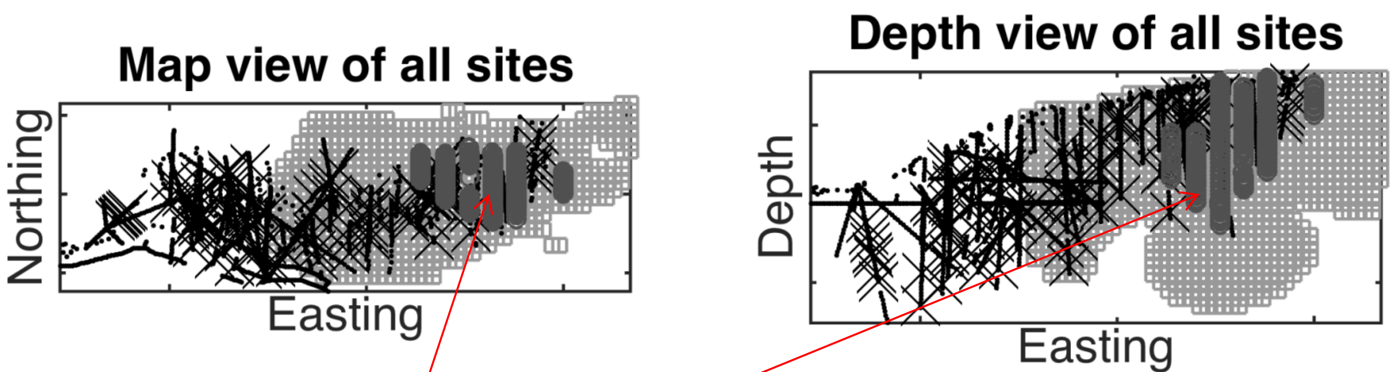




Planned boreholes.

Mining blocks. Some waste rock. Some high-grade.

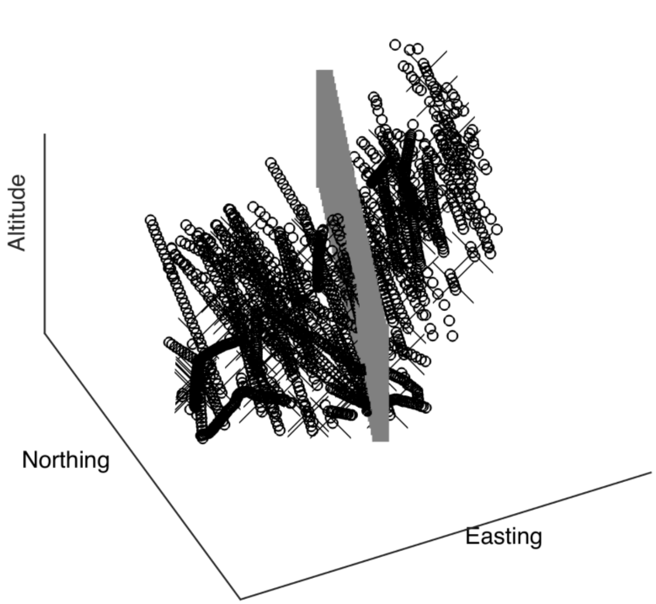
Information gathering

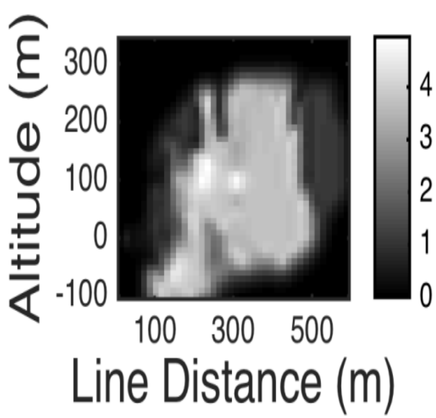


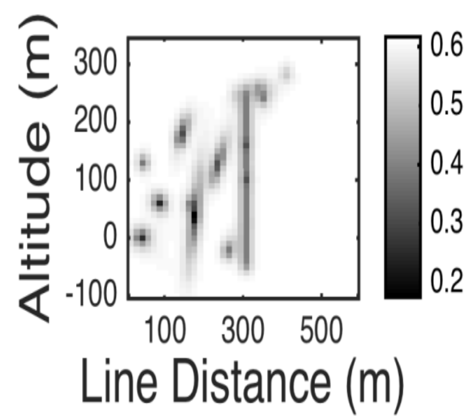
Planned boreholes.

- **Total test** : 265 measurements in 21 new boreholes.
- **Partial test**: Drilling and sampling data only in a subset of boreholes.
- **Perfect** testing (XRF: done in lab). **Imperfect** testing (XMET: handheld meter).

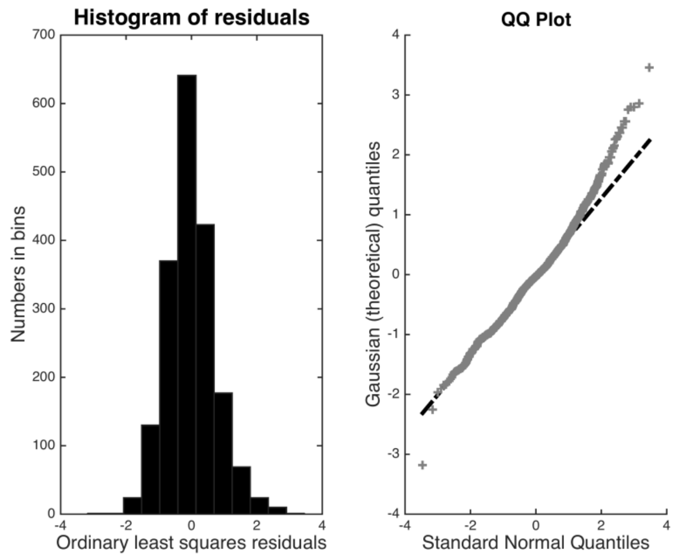
Prior model

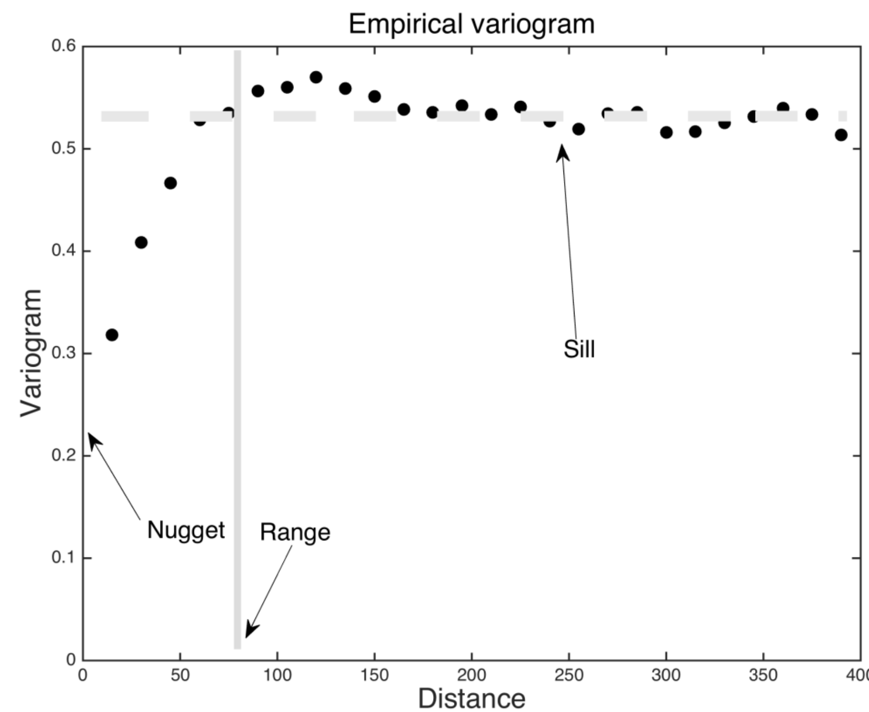

XMET
XRF
Altitude
Northing
Easting


Altitude (m)
300
200
100
0
-100
100
300
500
Line Distance (m)
4
3
2
1
0


Altitude (m)
300
200
100
0
-100
100
300
500
Line Distance (m)
0.6
0.5
0.4
0.3
0.2

Model


Histogram of residuals
Numbers in bins
Ordinary least squares residuals
QQ Plot
Gaussian (theoretical) quantiles
Standard Normal Quantiles


Empirical variogram
Variogram
Distance
Sill
Nugget
Range

Prior and likelihood model

Set from current data.

$$p(\boldsymbol{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = N(\boldsymbol{F}\boldsymbol{x}, \boldsymbol{T})$$

Defined by test (XRF, XMET).

Defined by design of boreholes.

$$y(\mathbf{s}) = x(\mathbf{s}),$$ -XRF data

$$y(\mathbf{s}) = x(\mathbf{s}) + N(0, \tau^2)$$ -XMET data

Results for design

Comparing designs over different criteria:

- All, 5 or 10 boreholes
- XRF or XMET data

- Geometric criterion,
- Variance reduction,
- Slope,
- Weight of mean,
- Entropy,
- VOI,

Results

Relative criterion	Geometric	Krig Std	Slope	Weight	Ent	VOI
5 XRF / All XMET	0.88	0.87	0.87	0.84	0.76	**0.39**
5 XRF/ All XRF	0.88	0.85	0.82	0.81	0.41	**0.27**
10 XRF / All XMET	0.94	0.96	0.97	0.95	**1.21**	0.94
10 XRF / All XRF	0.94	0.93	0.91	0.91	0.72	**0.64**
All XRF / All XMET	1	1.03	1.07	1.04	**1.68**	**1.45**
All XRF / All XRF	1	1	1	1	1	1

VOI

Weights set from block model(waste or ore).

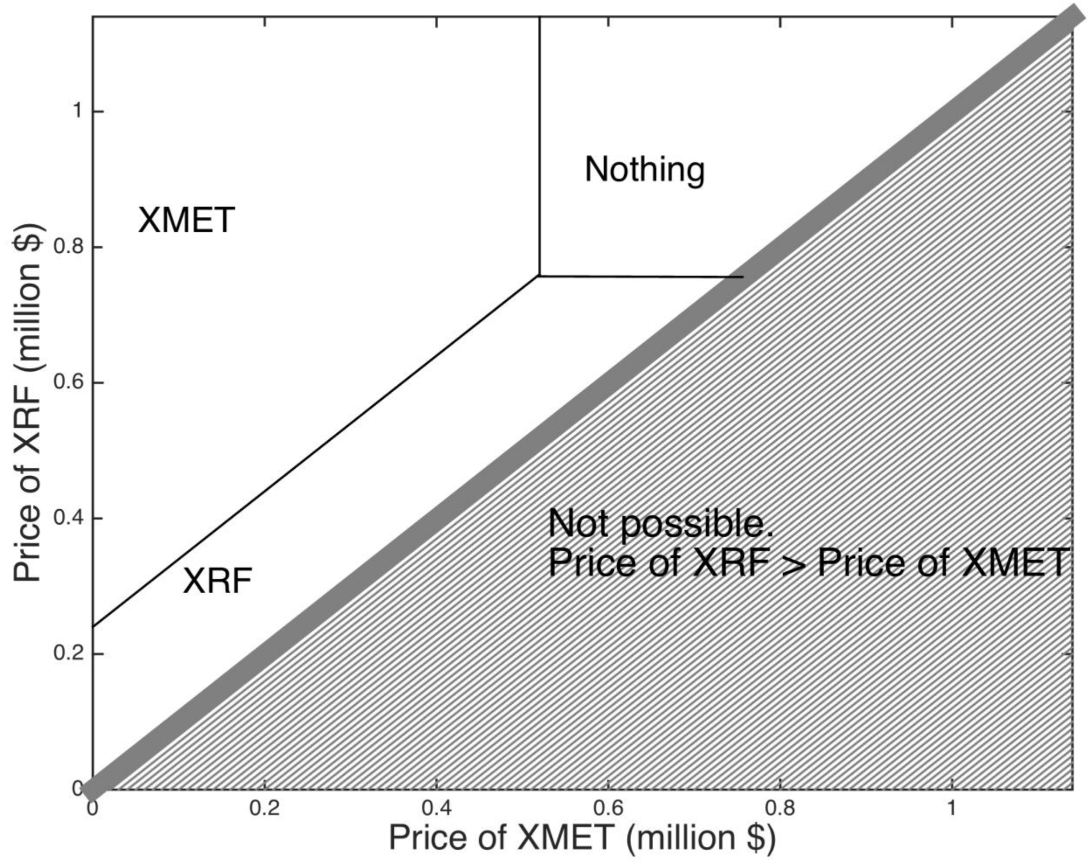
$$PV = \max\left\{E\left(\boldsymbol{w}^t\boldsymbol{x} - \mathrm{Cost}\right), 0\right\}$$

$$PoV\left(\boldsymbol{y}\right) = \int \max\left\{E\left(\boldsymbol{w}^t\boldsymbol{x} - \mathrm{Cost} \mid \boldsymbol{y}\right), 0\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

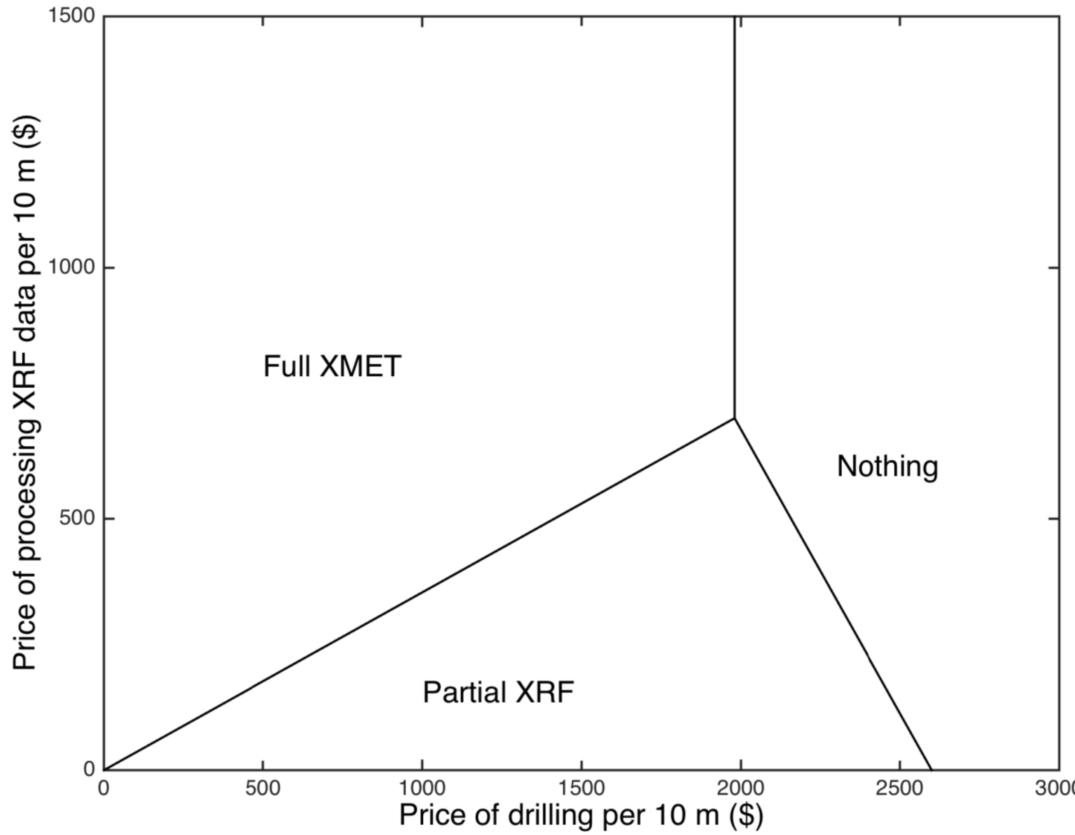
$$VOI\left(\boldsymbol{y}\right) = PoV\left(\boldsymbol{y}\right) - PV.$$

Analytical solution under the Gaussian modeling assumptions.

VOI : Decision regions XRF,XMET.



VOI : Decision regions, partial data.



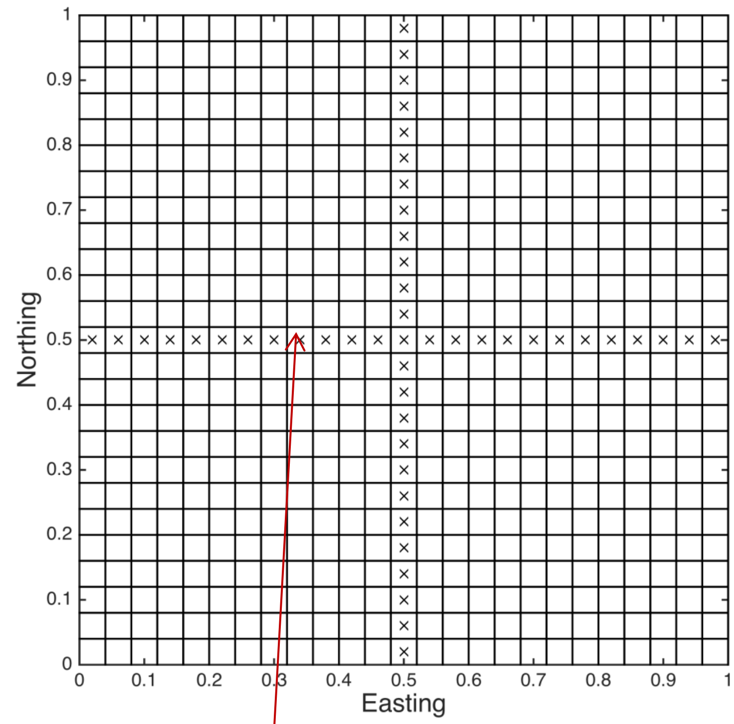
Take home from this exercise:

- Information connected to partial perfect testing can be less/more than total imperfect testing.
- Information criteria depend on design and data accuracy.
- Entropy appears to like perfect information.
- VOI can be connected with decisions and prices (not so easy for other criteria).

Norwegian wood - forestry example

Farmer must decide whether to harvest forest units, or not.

Another decision is whether to collect data before making these decisions.
If so, how and where should data be gathered.



Where to put survey lines for timber volumes information?
Typically partial, imperfect information.

Design

Design matrix:
Picks the measurement locations for a partial test.

$$\boldsymbol{x}_{\mathbb{K}} = \boldsymbol{F}\boldsymbol{x}$$

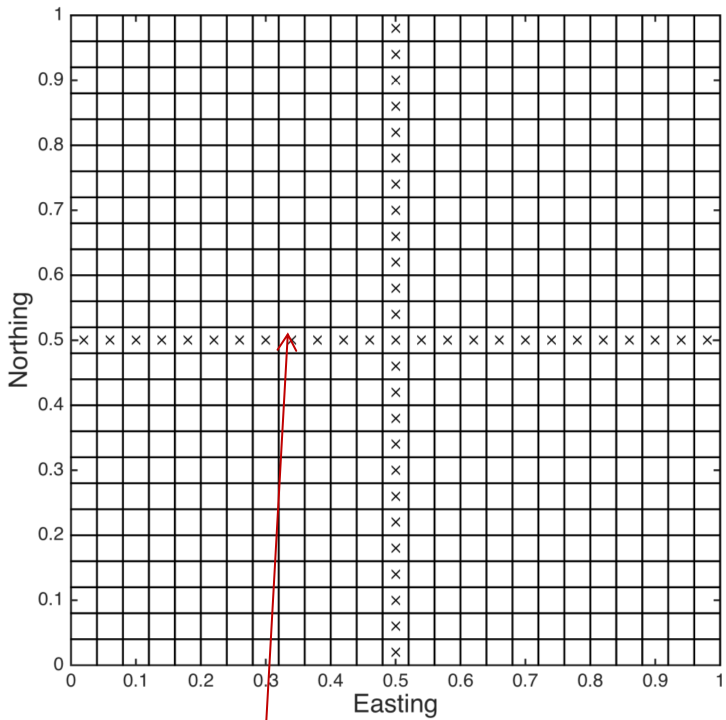
$$\boldsymbol{F} = \boldsymbol{I}$$

Total test.

$$\boldsymbol{y} = \boldsymbol{y}_{\mathbb{K}} = \boldsymbol{F}\boldsymbol{x} + N\left(\boldsymbol{0}, \tau^2\boldsymbol{I}\right)$$

$$p\left(\boldsymbol{y} \mid \boldsymbol{x}\right) = N\left(\boldsymbol{F}\boldsymbol{x}, \tau^2\boldsymbol{I}\right)$$

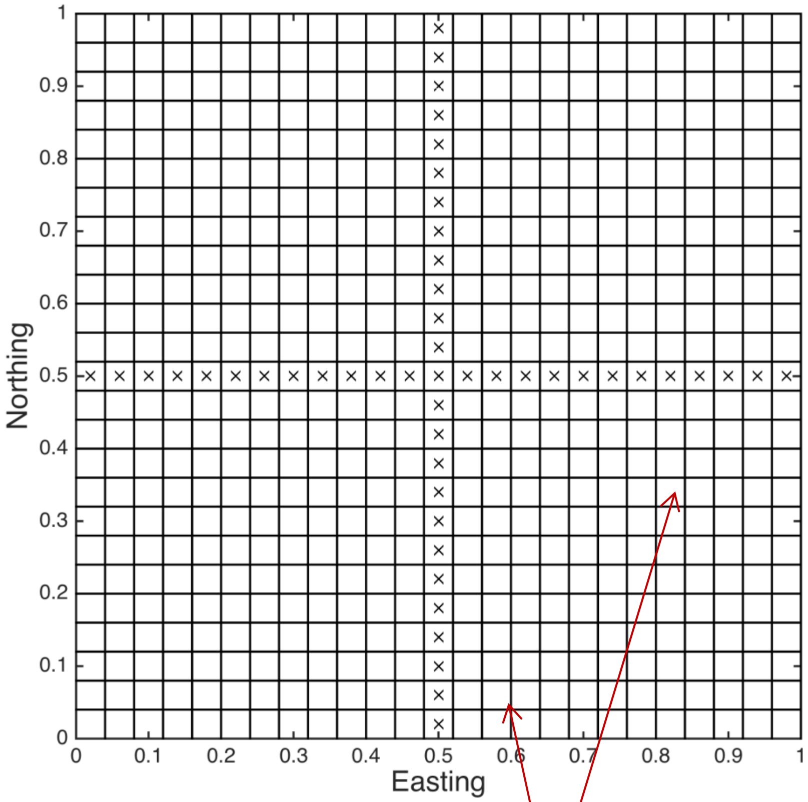
Example of imperfect test.



Where to put survey lines for timber volumes information?
Typically partial, imperfect information.

Three different designs

Three data designs:
- Total (all cells)
- Partial (all cells along center lines)
- Aggregate partial (sums along the two center lines).



Survey lines for timber volumes information?

Gaussian process for value

$$v = \sum_i v(x_i, a=1) = \sum_i x_i, \qquad v(x_i, a=0) = 0$$

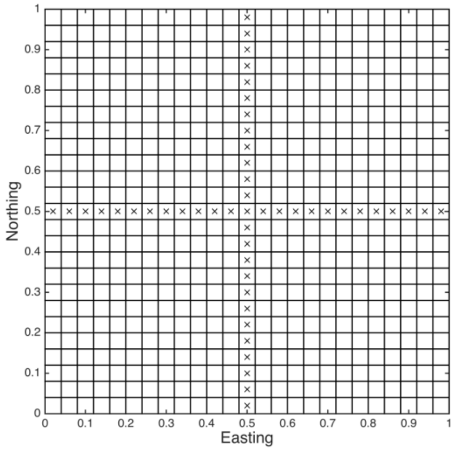
Global alternatives

$$v(x_i, a_i = 1) = x_i, \qquad v(x_i, a_i = 0) = 0$$

Local alternatives

Motivation, uncertainties on a grid - model profits directly (normalized for area).

Forest units.
Uncertainty is value in each cell.



Prior is Gaussian process

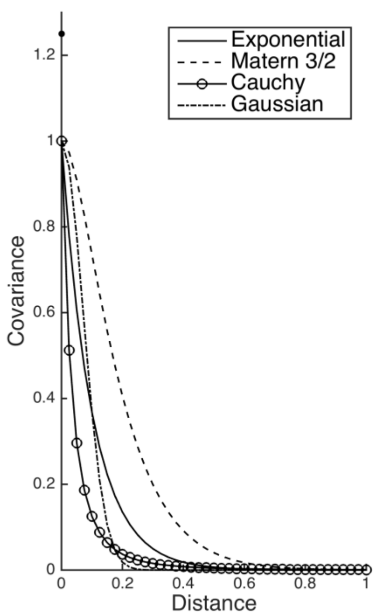
$$\boldsymbol{x} = \left(x_1, \ldots, x_n\right)$$

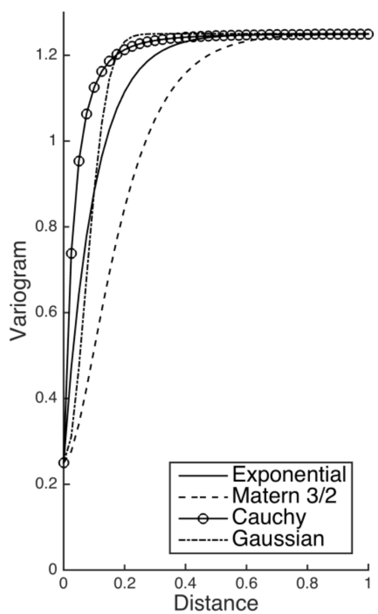
$$p\left(\boldsymbol{x}\right) = N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

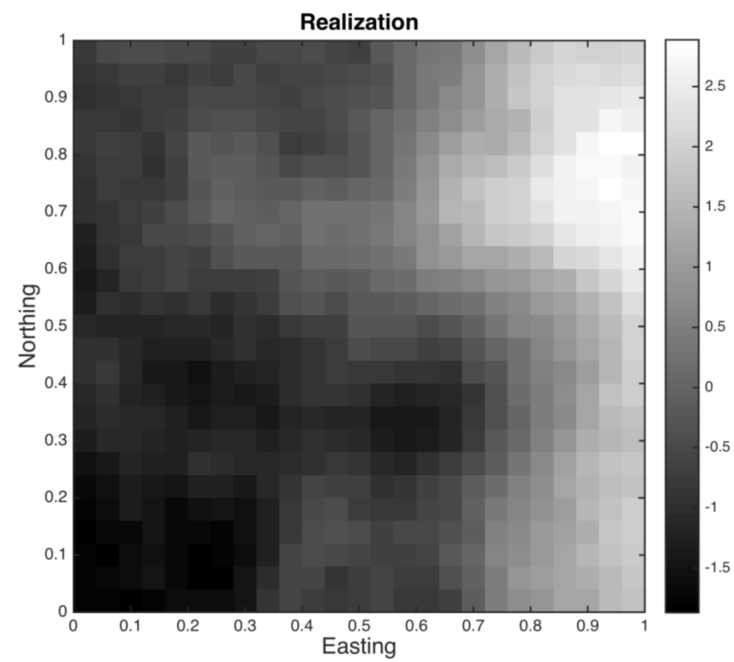
$$\boldsymbol{\mu} = \mathbf{1}\mu$$

$$\Sigma_{ij} = \Sigma\left(\mathbf{s}_i, \mathbf{s}_j; \theta\right)$$

Spatial dependence, Matern covariance.







Formulas for Gaussian models

$$p(\boldsymbol{x}) = N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

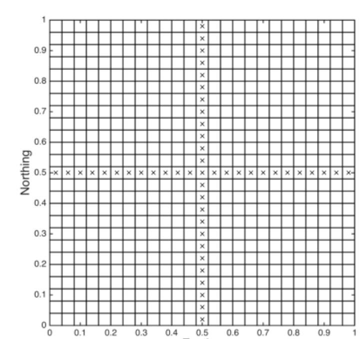
Prior for values.
Uncertainties are timber values.

$$\boldsymbol{y} = \boldsymbol{F}\boldsymbol{x} + N(\boldsymbol{0}, \tau^2 \boldsymbol{I})$$

$$p(\boldsymbol{y} \mid \boldsymbol{x}) = N(\boldsymbol{F}\boldsymbol{x}, \tau^2 \boldsymbol{I})$$

Likelihood,
design matrix, picks
data locations.

Motivation, uncertainties on a grid – forest units.



Conditioning – Gaussian models

$$E(\boldsymbol{x} \mid \boldsymbol{y}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}\boldsymbol{F}^t\left(\tau^2\boldsymbol{I} + \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t\right)^{-1}(\boldsymbol{y} - \boldsymbol{F}\boldsymbol{\mu})$$

$$Var(\boldsymbol{x} \mid \boldsymbol{y}) = \boldsymbol{\Sigma} - \boldsymbol{R}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma}\boldsymbol{F}^t\left(\tau^2\boldsymbol{I} + \boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t\right)^{-1}\boldsymbol{F}\boldsymbol{\Sigma}$$

$$r_i = \sqrt{R_{ii}}$$

Prediction, Kriging.

VOI – Gaussian models

Low flexibility:
Must select all units, or none.

$$PV = \max\left\{0, E\left(\sum_{i=1}^{n} x_i\right)\right\} = \max\left\{0, \sum_{i=1}^{n} \mu_i\right\}$$

Value decouples to sum.

$$\boldsymbol{R} = \boldsymbol{\Sigma F}^t\left(\tau^2 \boldsymbol{I} + \boldsymbol{F\Sigma F}^t\right)^{-1} \boldsymbol{F\Sigma}$$

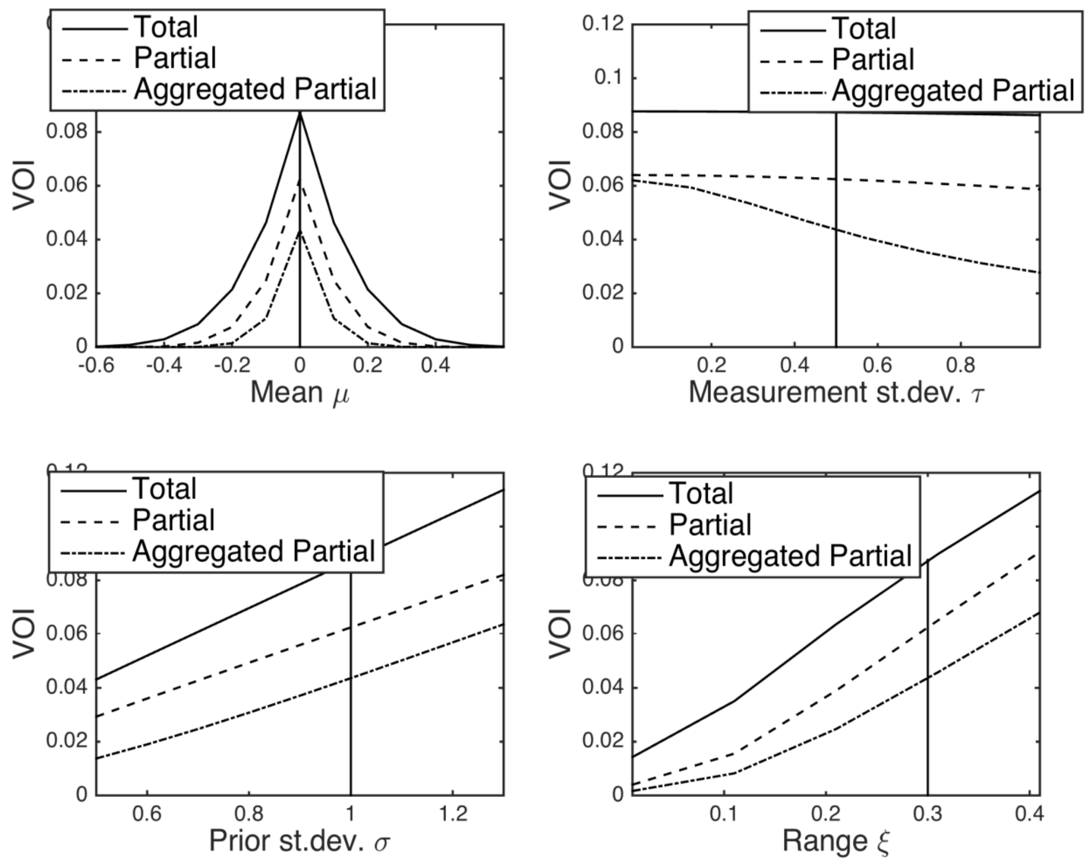
$$\mu_w = \sum_{i=1}^{n} \mu_i \qquad r_w^2 = \sum_{k=1}^{n}\sum_{l=1}^{n} R_{kl}$$

$$PoV(\boldsymbol{y}) = \int \max\left\{0, E\left(\sum_{i=1}^{n} x_i \mid \boldsymbol{y}\right)\right\} p(\boldsymbol{y})\,d\boldsymbol{y} = \mu_w \Phi\left(\mu_w / r_w\right) + r_w \phi\left(\mu_w / r_w\right)$$

$\phi(z), \Phi(z)$ standard Gaussian density and cumulative function

Results - Forestry example

Low flexibility:
Must select all units, or none.



Total: all cells. Partial: Every cell along center lines. Aggregated partial: sums along center lines.
(Results are normalized for area).

Insight in VOI from this example

- Total test does not necessarily give much higher VOI than a partial test. It depends on the spatial design of experiment as well as the prior model (mean and dependence).
- VOI increases with larger dependence in spatial uncertainties.
- VOI is largest when we are most indifferent in prior (mean near 0 and large prior uncertainty.
- VOI increases with higher accuracy of measurements.

Project : Norwegian wood

- Consider profits from timber in a forest split in many units. Profits are modeled as a Gaussian random field represented on a 25 x 25 grid for the 625 units. The mean is m=0 at all cells, the covariance is exponential with st dev r=1 and correlation range r=40. Use the code to draw a random realization of this Gaussian process.

- One can gather imperfect data at 100 random design locations, giving unbiased profit measurements, and independent error with st dev 0.5. Use the code to draw a random dataset.

- Compute the Kriging prediction and the associated variances. See code.

- Compute the VOI (using the same data design) of the decision situation where the farmer harvest all units or none.

- Compare the VOI results (using the same data design) for different prior mean, variances, correlation ranges and measurement noise terms.