Value of Information

Spatial Statistics, Design of Experiments and Value of Information Analysis

Jo Eidsvik

Jo.Eidsvik@ntnu.no

Plan for course

Time	Topic
Lecture 1	Introduction and motivating examples
	Elementary decision analysis and the value of information
	Multivariate statistical modeling, dependence, graphs
	Value of information analysis for dependent models
Lecture 2	Re-cap of VOI and statistical dependence
	Spatial statistics, spatial design of experiments
	Value of information analysis in spatial decision situations
	Examples of value of information analysis in Earth sciences
Lecture 3	Computational aspects of VOI analysis, approximate calculations
	Sequential information gathering
	Examples from Earth sciences

Every day: Small exercises.

Bayesian model

- All the currently available information is contained in the prior model for the variables:

$$p(\boldsymbol{x})$$

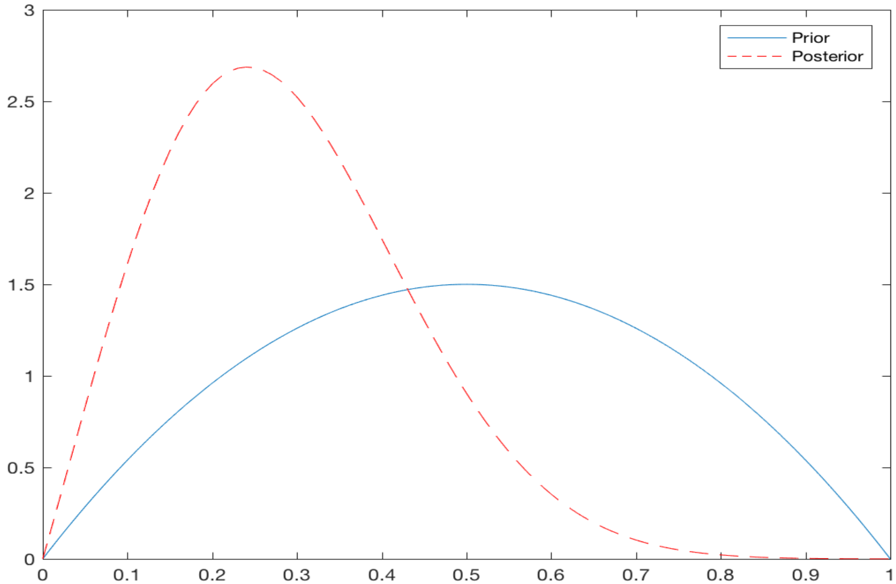
- New data (and the data gathering scheme) is represented by a likelihood model:

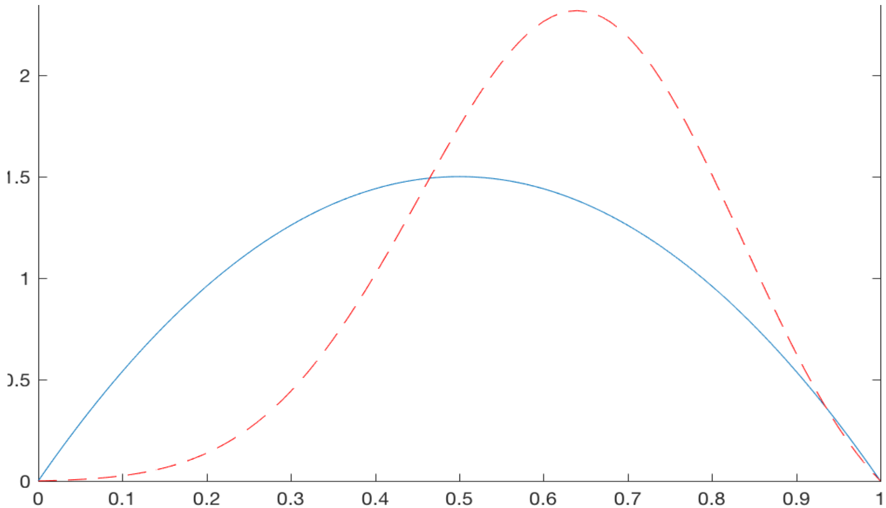
$$p(\boldsymbol{y} \mid \boldsymbol{x})$$

- If we collect data, the model is updated to the posterior, conditional on the new observations:

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

Bayesian updating





$$p(\boldsymbol{x})$$

$$p(\boldsymbol{x} \mid \boldsymbol{y})$$

- What data is valuable?
- Study the **expected effect of data,** before it is collected.
- We gather data not only to reduce uncertainty, but to make better **decisions**. We have a goal, a clear question we want to answer.

Decision analysis

- Uncertain variables: $\boldsymbol{x} = \left(x_1, \ldots, x_n \right)$

- Alternatives (Where? How? When?) $\boldsymbol{a} = \left(a_1, \ldots, a_N \right)$
- Value function is revenues, subtracted costs. $v\left(\boldsymbol{x}, \boldsymbol{a} \right)$

- Risk neutral decision maker will **maximize expected value**:

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

Value of information (VOI)

Prior value:

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

Posterior value:

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

VOI = Expected posterior value – Prior value

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

$\boldsymbol{x}$ - Uncertainties

$\boldsymbol{a}$ - Alternatives

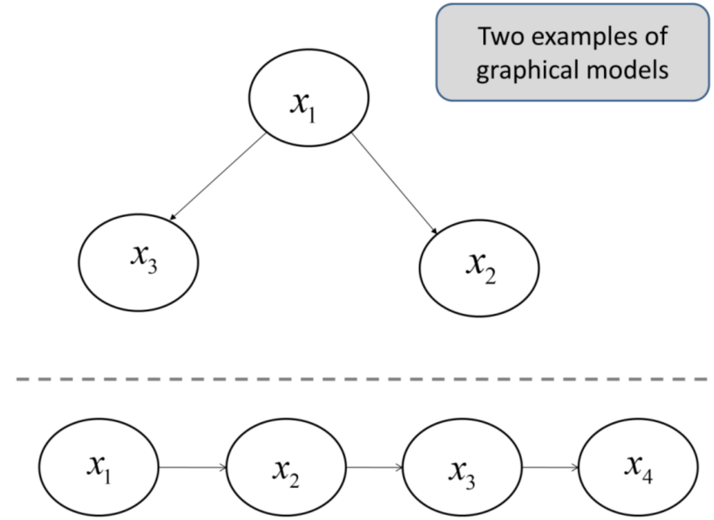
$v\left(\boldsymbol{x},\boldsymbol{a}\right)$ - Value function

$\boldsymbol{y}$ - Data

Information gathering

	Perfect	Imperfect
Total	Exact observations are gathered for all locations. $$\boldsymbol{y} = \boldsymbol{x}$$	Noisy observations are gathered for all locations. $$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{\varepsilon}$$
Partial	Exact observations are gathered at some locations. $$\boldsymbol{y}_{\mathbb{K}} = \boldsymbol{x}_{\mathbb{K}}, \quad \mathbb{K} \text{ subset}$$	Noisy observations are gathered at some locations $$\boldsymbol{y}_{\mathbb{K}} = \boldsymbol{x}_{\mathbb{K}} + \boldsymbol{\varepsilon}_{\mathbb{K}}, \quad \mathbb{K} \text{ subset}$$

Markov chains



Markov chains are special graphs, defined by initial probabilities and transition matrices.

$$p(\mathbf{x}) = p(x_1, x_2, \ldots, x_n) = p(x_1)\, p(x_2 \mid x_1) \ldots p(x_n \mid x_{n-1})$$

$$p(x_1 = k), \qquad k = 1, \ldots, d$$

$$p(x_{i+1} = l \mid x_i = k) = P(k, l), \qquad k, l = 1, \ldots, d$$

$d = 2$

$$P = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix} \qquad P = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \qquad P = \begin{bmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{bmatrix} \qquad P = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix}$$

Independence

Absorbing

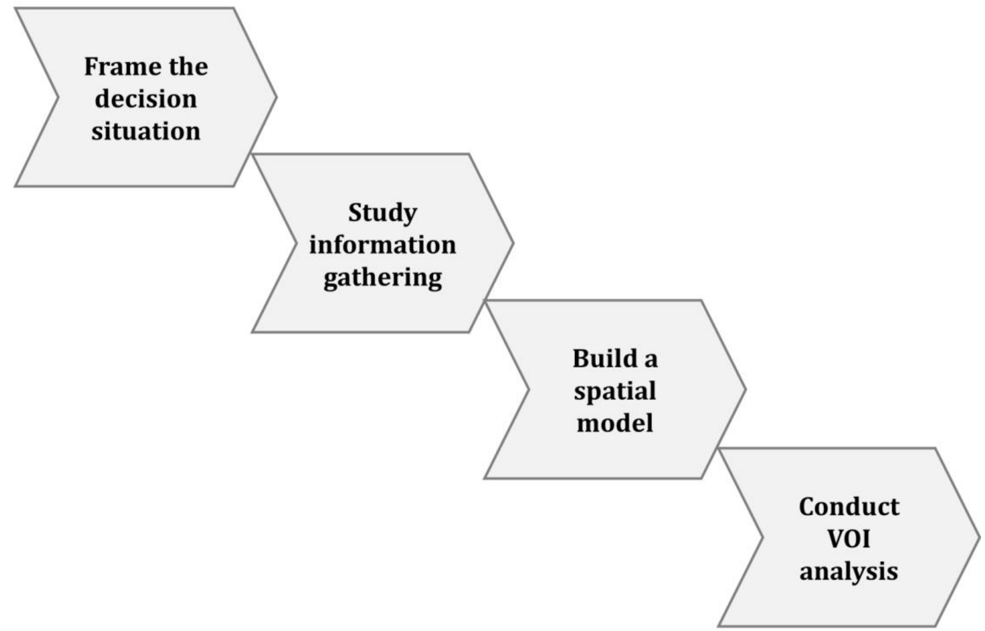
Avalanche decisions and sensors

Suppose that parts along a road are at risk of **avalanche**.
- One can remove risk by clearing roads, at a cost.
- Otherwise, the repair cost depends on the **unknown risk class: 1) low, 2) high**.

Data, sensor at a particular location, can help classify the risk class and hence improve the decisions made regarding cleaning / wait and see.

VOI workflow

- Clear entire road up front (fixed cost), or wait and see (uncertain cost at each location).
- Gather information by sensor at one location, perfect information about risk class at that location.
- Model is a Markov chain with increasing probability of high risk for later indeces (altitude).
- VOI analysis done by Markov chain calculations. Conducted for all possible sensor locations.



Avalanche decisions - risk analysis

n=50 identified locations along railroad track, at increasing altitude and risk of **avalanche**.
One can remove risk entirely by cost 100 000.
If it is not removed, the repair cost, at each location, depends on the unknown risk class:

$$C_j, \qquad j \in \{1,2\},$$

$$\mathrm{C}_1 = 0, \mathrm{C}_2 = 5000,$$

Decision maker must choose whether to

- **i)** **clean tracks** up front, with fixed price.
- **ii)** **wait and see**, with the uncertain price at each location.

The decision is based on the minimization of expected costs.

Prior value:

$$PV = \max\left\{-100000, -5000\sum_{i=1}^{50} p(x_i = 2)\right\}$$

Clean up front — Expected value when wait and see.

Markovian model for risk of avalanche

Risk tends to start in lower class (1), and then move to higher class (2).
If risk class 2 is reached, it will stay there until location 50 (absorbing state).

$$x_i \in \{1,2\}, \qquad i = 1,\dots,50,$$

$$p(x_1 = 1) = 0.99,$$

$$p(x_1 = 2) = 0.01,$$

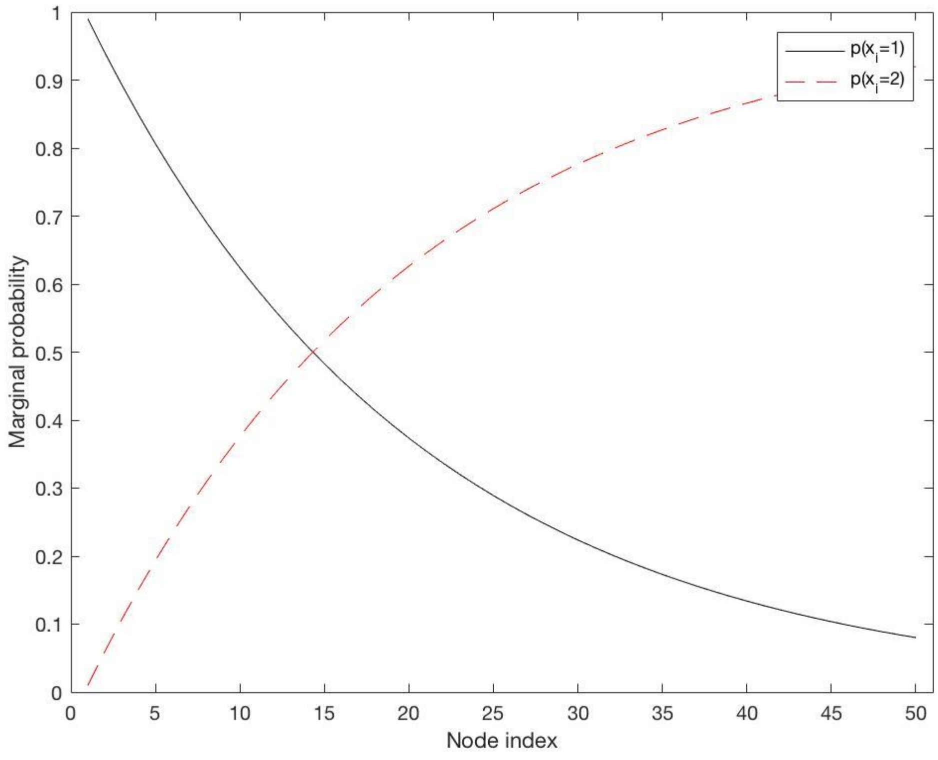
$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0 & 1 \end{bmatrix}$$

Absorbing!

Results – marginals

$$p\left(x_i = l\right) = \sum_{k=1}^{2} p\left(x_{i-1} = k\right) p(x_i = l \mid x_{i-1} = k), \qquad l = 1, 2, \qquad i = 1, \ldots, n$$

$$p\left(x_i = 1\right) = p \cdot 0.99^{i-1} \qquad i = 1, \ldots, n$$



Sensor – perfect risk information at one location

- Install a sensor at one location, getting perfect information at that node.
- Compute conditional probabilities.

$$p\left(x_i = k \mid x_j = l\right), \qquad i = 1, \ldots, 50$$

Results – conditionals (forward)

$$p\left(x_i = k \mid x_j = l\right) = \sum_{q=1}^{2} p\left(x_i = k, x_{i-1} = q \mid x_j = l\right) = \sum_{q=1}^{2} P\left(q,k\right) p\left(x_{i-1} = q \mid x_j = l\right)$$

$$p\left(x_i = 1 \mid x_j = 1\right) = 0.99^{i-j} \qquad i \geq j,$$

$$p\left(x_i = 2 \mid x_j = 2\right) = 1 \qquad i \geq j$$

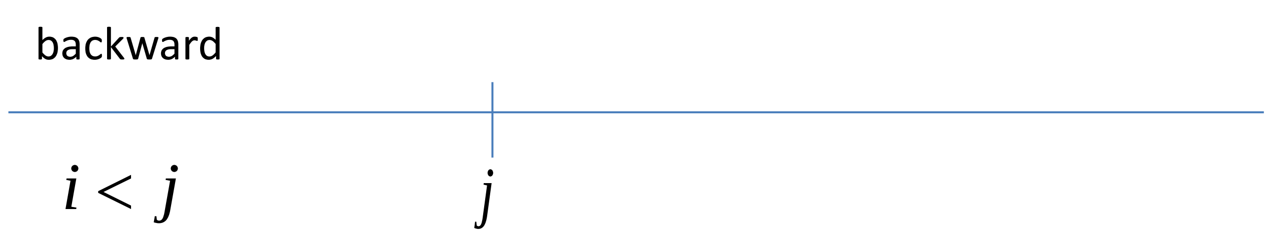
Results – conditionals (backward)

$$p\left(x_i = k \mid x_{i+1},\ldots,x_n\right) = p\left(x_i = k \mid x_{i+1} = l\right) = \frac{p\left(x_i = k, x_{i+1} = l\right)}{p\left(x_{i+1} = l\right)} = \frac{P\left(k,l\right) p\left(x_i = k\right)}{p\left(x_{i+1} = l\right)}$$

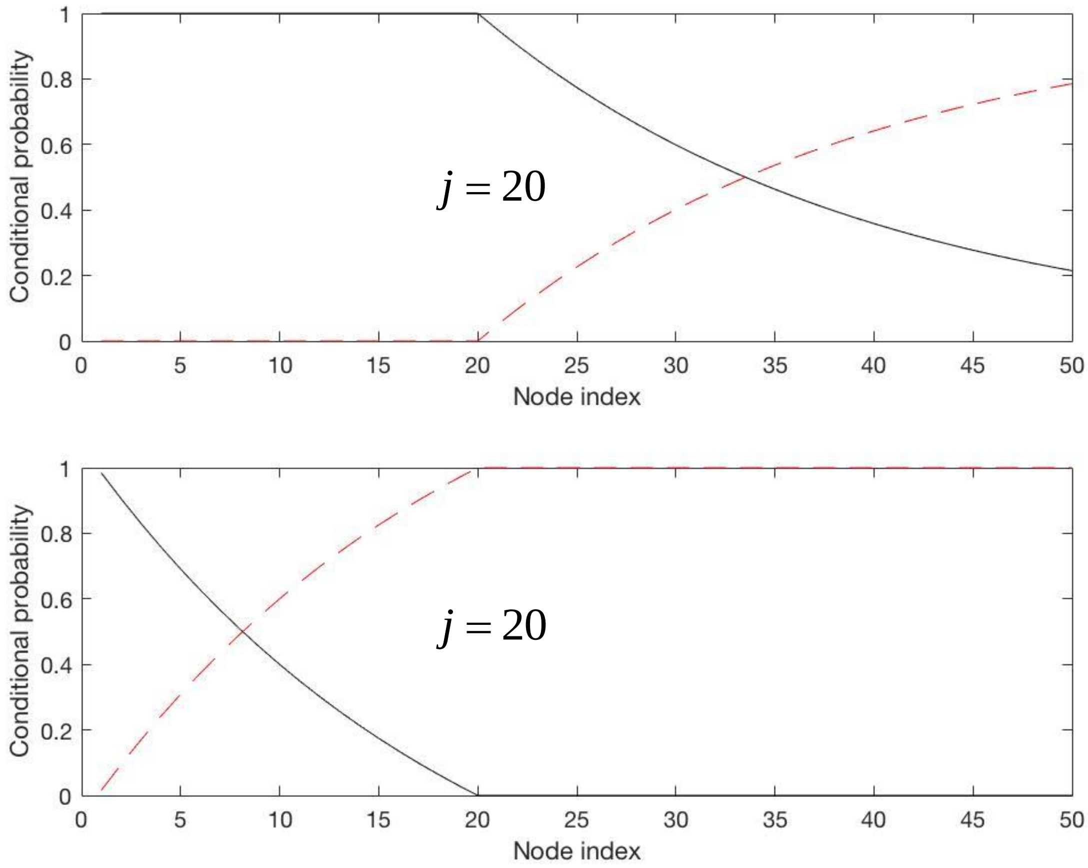
$$p\left(x_i = k \mid x_j = l\right) = \sum_{q=1}^{2} p\left(x_i = k \mid x_{i+1} = q\right) p\left(x_{i+1} = q \mid x_j = l\right)$$

$$p\left(x_i = 1 \mid x_j = 1\right) = 1 \qquad i < j,$$

$$p\left(x_i = 2 \mid x_j = 2\right) = \frac{p\left(x_i = 2\right)}{p\left(x_j = 2\right)} \qquad i < j$$



Results – conditional probabilities



Learning risk of avalanche

- Plan to install a sensor at one location, getting perfect information at that location.

$$j \in \{1, \ldots, 50\}$$

- Compute the posterior value, with sensor location at one location.
  Compute the VOI.
- What is the optimal sensor location, if the goal is to improve risk decisions?

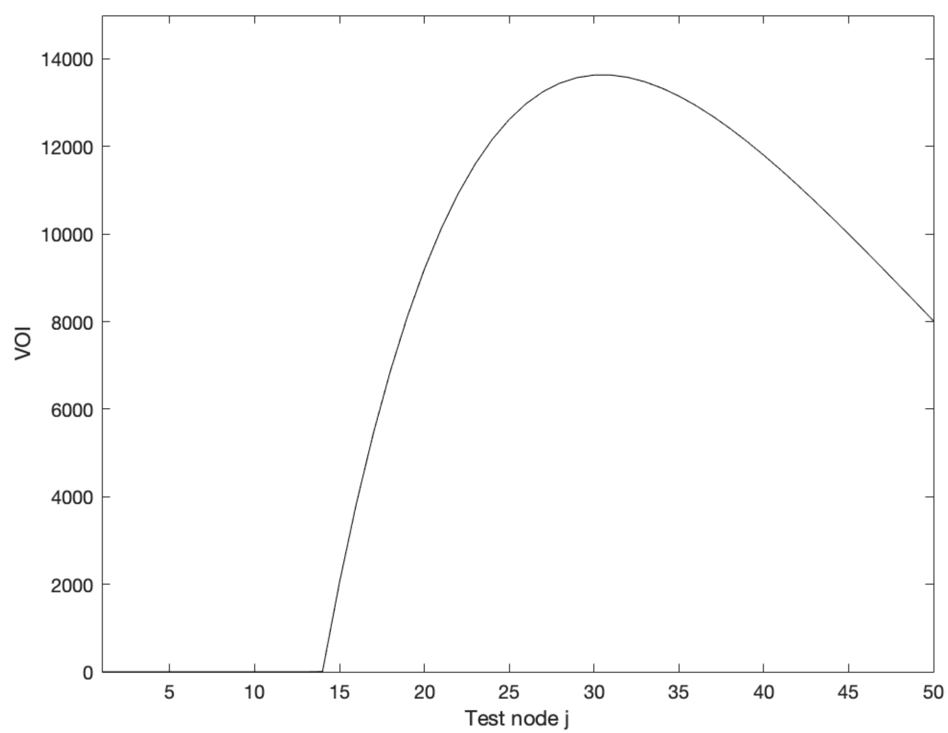
Results – VOI

$$PV = \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2\right)\right\}$$

$$PoV\left(x_j\right) = \sum_{k=1}^{2} \max\left\{-100000, -5000\sum_{i=1}^{50} p\left(x_i = 2 \mid x_j = k\right)\right\} p\left(x_j = k\right)$$

$$VOI\left(x_j\right) = PoV\left(x_j\right) - PV$$

Best location near *j=30*.
The VOI is about 13000



Hands on - Avalanche risk

n=50 identified locations along railroad track, at risk of **avalanche**.

Decision maker must choose whether to

i) **clean tracks** up front, with fixed price.

ii) **wait and see**, with the uncertain price.

$$PV = \max\left\{-100000, -C_2 \sum_{i=1}^{50} p\left(x_i = 2\right)\right\}$$

$$PoV\left(x_j\right) = \sum_{k=1}^{2} \max\left\{-100000, -C_2 \sum_{i=1}^{50} p\left(x_i = 2 \mid x_j = k\right)\right\} p\left(x_j = k\right)$$

Is the VOI sensitive to C_2*=5000 (4000, 6000)* ?

Is the optimal sensor location sensitive to C_2*=5000 (4000, 6000)* ?

Implement in Python, R or Matlab.

$$x_i \in \{1, 2\}, \qquad i = 1, \ldots, 50,$$

$$p\left(x_1 = 1\right) = 0.99,$$

$$p\left(x_1 = 2\right) = 0.01,$$

$$P = \begin{bmatrix} 0.95 & 0.05 \\ 0 & 1 \end{bmatrix}$$

Gaussian distribution and VOI

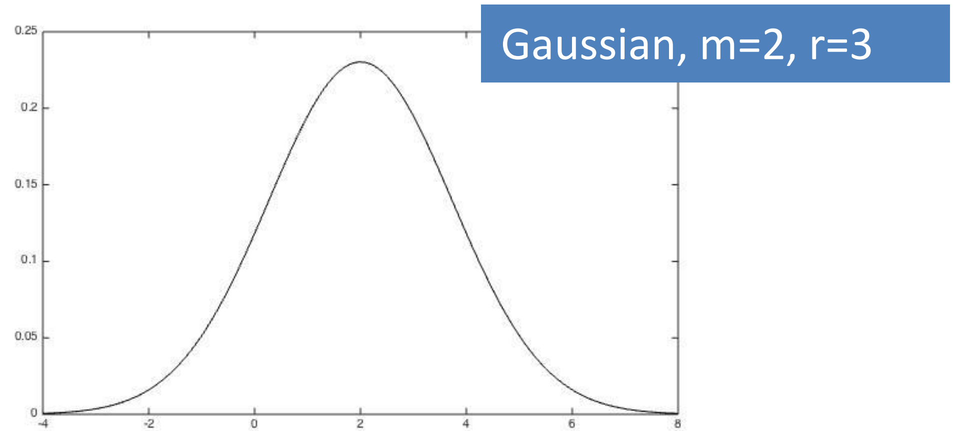
Main topic today is computing the VOI in spatial Gaussian linear models.

- The spatial random variables are assumed to be Gaussian distributed
- Make good **spatial designs**.
- The optimal design depends on the model and the decision situation.

Gaussian distribution is very common and analytical properties are available.

Gaussian profit

$$p(x) = \frac{1}{\sqrt{2\pi r^2}} \exp\left(-\frac{(x-m)^2}{2r^2} \right)$$



Uncertain profit of a project is Gaussian distributed.

VOI for Gaussian project profits

Uncertain project profit is Gaussian distributed.
Invest or not?
The decision maker asks a clairvoyant for perfect information, if the VOI is larger than her price.

$$VOI(x) = \text{PosteriorValue}(x) - \text{PriorValue}$$

$$PV = \max\{0, E(x)\}, \qquad E(x) = m$$

$$PoV(x) = E(\max\{0, x\}) = \int \max\{0, x\}\, p(x)\, dx$$

Posterior value of perfect information

Result:

$$E\left(\max\{0,x\}\right)=\int \max\{0,x\}\,p(x)\,dx=\int_{0}^{\infty} x p(x)\,dx=\int_{-m/r}^{\infty}(m+rz)\phi(z)\,dz$$

$$=m\int_{-m/r}^{\infty}\phi(z)\,dz+r\int_{-m/r}^{\infty}z\phi(z)\,dz=m\left(1-\Phi\left(-m/r\right)\right)+r\phi\left(-m/r\right)$$

$$=m\Phi\left(m/r\right)+r\phi\left(m/r\right),$$

Gaussian cdf

Gaussian pdf

VOI for Gaussian

Gaussian cdf

Gaussian pdf

Result:

$$VOI(x) = m\Phi\left(\frac{m}{r}\right) + r\phi\left(\frac{m}{r}\right) - \max\{0, m\}$$

The analytical form facilitates computing, and eases the study of VOI properties as a function of the parameters.

VOI for Gaussian

Gaussian cdf

Gaussian pdf

Result:

$$VOI(x) = m\Phi\left(\frac{m}{r}\right) + r\phi\left(\frac{m}{r}\right) - \max\{0, m\}$$

i) The VOI is largest with mean at 0, most difficult to make decision.

ii) If $m = 0$,

$$VOI(x) = r\phi(0) = \frac{r}{\sqrt{2\pi}}$$

More uncertain -> information more valuable.

Multivariate Gaussian and VOI

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

Standard bivariate
Gaussian:

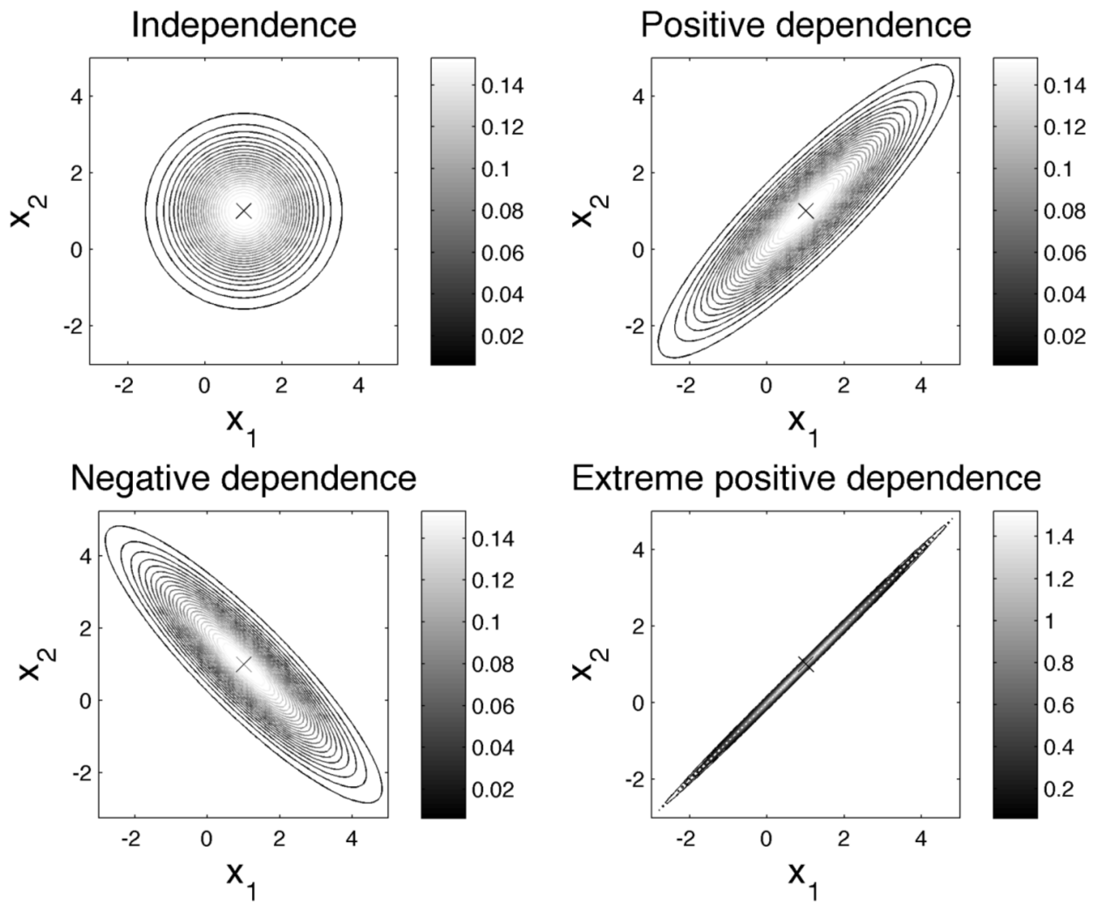
$$p(\boldsymbol{x}) = N(\mathbf{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Two-project example - Gaussian

Two correlated projects with uncertain profits.

Decision maker considers investing in project(s).

The prior distribution is the bivariate Gaussian.



Gaussian projects example

- **Alternatives**
  - Do not invest in project 1 (a1=0) - Invest in project 1 (a1=1)
  - Do not invest in project 2 (a2=0) - Invest in project 1 (a2=1)
  - Decision maker is free to select both, if profitable: Four sets of alternatives.
- **Uncertainty** (random variable)
  - Profits are *bivariate Gaussian*.
    Assume mean 0, variance 1 and fixed correlation.
- **Value** decouples to sum of profits, if positive.
- **Information gathering**
  - Report can be written about one project (assume perfect information).
  - Report can be written about both projects (assume imperfect information).

Gaussian projects example

$$\boldsymbol{x} = (x_1, x_2)$$

Prior model for profits: $p(\boldsymbol{x}) = N(\boldsymbol{0}, \boldsymbol{\Sigma}),$ $\boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$

$$PV = \sum_{i=1}^{2} \max\left\{0, E(x_i)\right\} = 0 + 0 = 0$$

$$PoV(\boldsymbol{y}) = \sum_{i=1}^{2} \int \max\left\{0, E(x_i \mid \boldsymbol{y})\right\} p(\boldsymbol{y}) d\boldsymbol{y}$$

$$\text{VOI}(\boldsymbol{y}) = PoV(\boldsymbol{y}) - PV$$

Gaussian projects example

$$PV = \sum_{i=1}^{2} \max\left\{0, E\left(x_i\right)\right\} = 0 + 0 = 0$$

$$PoV\left(\boldsymbol{y}\right) = \sum_{i=1}^{2} \int \max\left\{0, E\left(x_i \mid \boldsymbol{y}\right)\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

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

Must solve the integral expression!

Need conditonal expectation!

Need marginal for data!

Perfect information about 1 project

$$y = x_1$$

$$p(x_1) = N(0,1)$$

$$E(x_1 \mid x_1) = x_1$$

$$E(x_2 \mid x_1) = \rho x_1$$

Get information about second project because of correlation!

$$Var(x_1 \mid x_1) = 0, Var(x_2 \mid x_1) = 1 - \rho^2$$

$$PoV(x_1) = \int_0^\infty x_1 p(x_1) dx_1 + \int_0^\infty |\rho| x_1 p(x_1) dx_1$$

$$= \frac{(1+|\rho|)}{\sqrt{2\pi}}$$

Imperfect information, both projects

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

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

$$E\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \boldsymbol{\Sigma} \boldsymbol{C}^{-1} \boldsymbol{y}$$

Reduction in variances large, VOI is large.

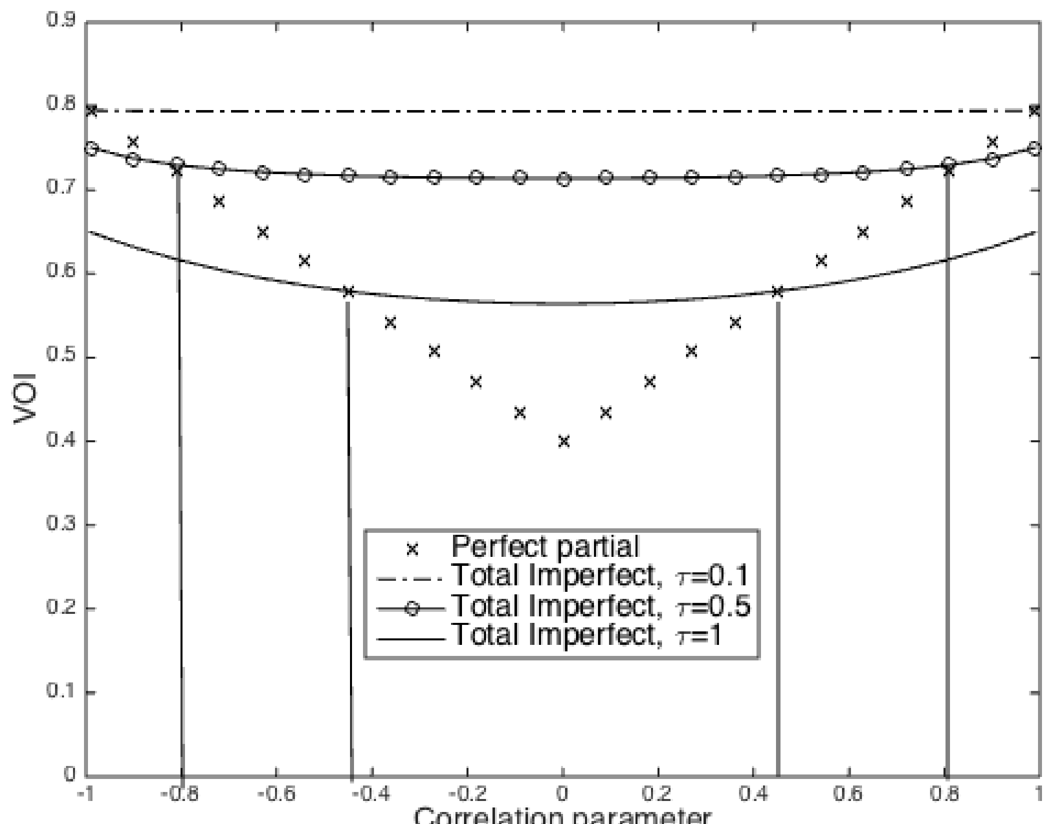
$$Var\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \boldsymbol{\Sigma} - \boldsymbol{R}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma} \boldsymbol{C}^{-1} \boldsymbol{\Sigma}$$

$$PoV\left(\boldsymbol{y}\right) = \sum_{i=1}^{2} \int \max\left\{0, E\left(x_i \mid \boldsymbol{y}\right)\right\} p\left(\boldsymbol{y}\right) d\boldsymbol{y} = \left(\sqrt{R_{1,1}} + \sqrt{R_{2,2}}\right) \Big/ \sqrt{2\pi}$$

Gaussian projects results

$$PoV(\boldsymbol{y}) = \left(\sqrt{R_{1,1}} + \sqrt{R_{2,2}}\right) \Big/ \sqrt{2\pi}, \qquad \boldsymbol{R} = \boldsymbol{\Sigma}\boldsymbol{C}^{-1}\boldsymbol{\Sigma}$$

$$PoV(x_1) = \left(1 + |\rho|\right) \Big/ \sqrt{2\pi}$$



Insight from Gaussian projects

Dependence matters – the more correlation, the larger VOI.
The relative increase is very clear for partial information. It is also larger when there is more measurement noise. (With perfect total information, dependence does not matter.)

Decision maker must compare the VOI with the price of information, or a budget.
VOI is often used to compare different possible designs.

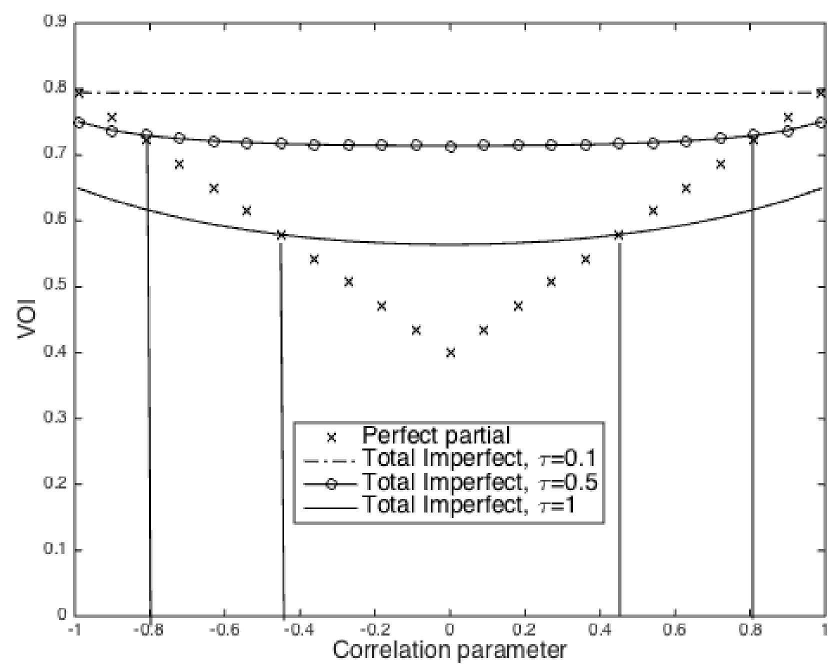
Exercise : Two Gaussian projects

Consider the bivariate Gaussian projects example, with prior mean 0 and variance 1, correlation 0.7 (and 0.1) and measurement noise st dev 1.

- Study the *decision regions* for no testing, partial perfect or total imperfect testing:

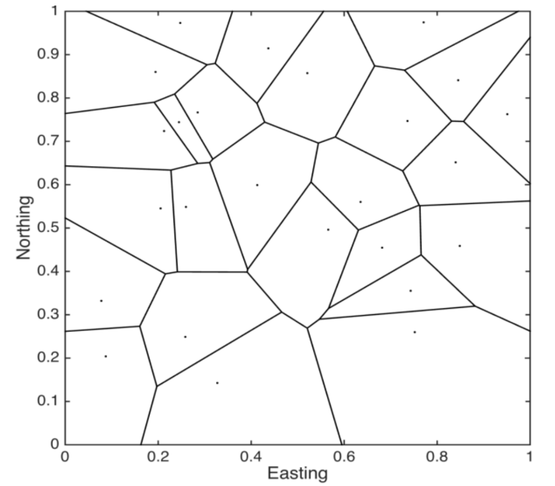
$$\arg\max\left\{VOI_{1,2}-\left(P_{1,2,imp}\right),VOI_1-P_{1,perf},0\right\}$$

Decision regions are visual plots, with the price of one perfect on the x-axis, and the price of two imperfect on the y-axis.



Joint Gaussian pdf – spatial field

Gaussian process (for some discretization of space):



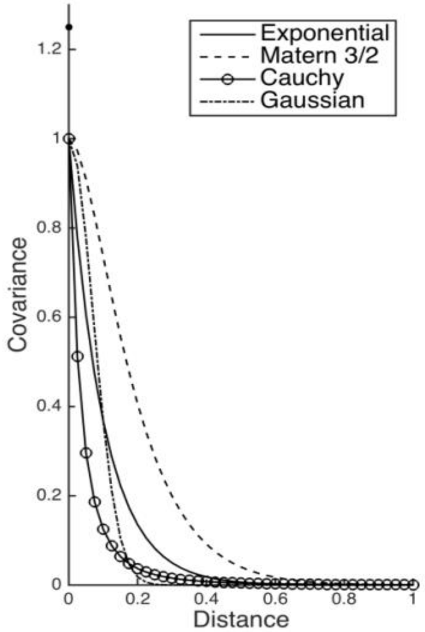
$$p(\boldsymbol{x}) = N(\boldsymbol{0}, \boldsymbol{\Sigma}), \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \dots & \Sigma_{1n} \\ \Sigma_{21} & \Sigma_{22} & .. & \Sigma_{21} \\ .. & .. & .. & .. \\ \Sigma_{n1} & \Sigma_{n2} & .. & \Sigma_{nn} \end{pmatrix}$$

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

Spatial covariance functions

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

$$\Sigma_{ij} = \Sigma\left(\left|\mathbf{s}_i - \mathbf{s}_j\right|\right) = \Sigma\left(\left|\mathbf{t}\right|\right)$$



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

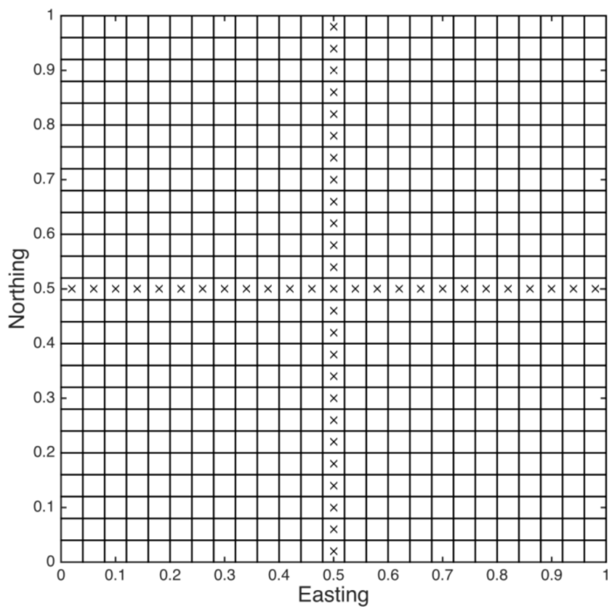
Model for data

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

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

Design matrix: $\boldsymbol{F}$

Example: Data collected along center lines.



Gaussian posterior model

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

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

Goal is:

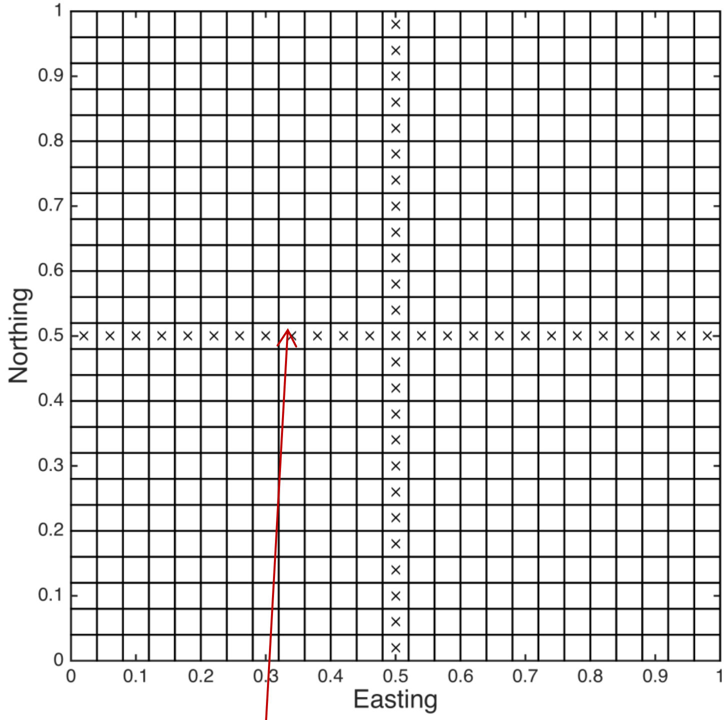
$$p(\boldsymbol{x} \mid \boldsymbol{y}) = \frac{p(\boldsymbol{y} \mid \boldsymbol{x})\, p(\boldsymbol{x})}{p(\boldsymbol{y})}$$

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

Norwegian wood - forestry example

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

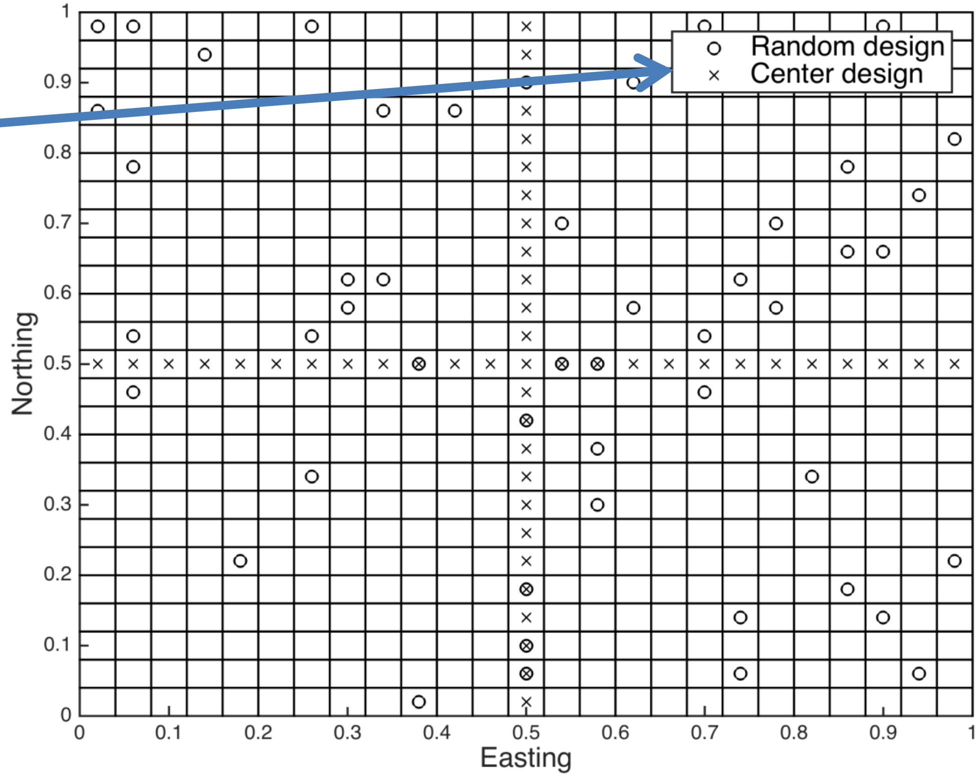
Another decision is whether to collect data before making these decisions.
If so, how and where should data be gathered.



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

Norwegian wood - posterior

Design matrix: $\boldsymbol{F}$



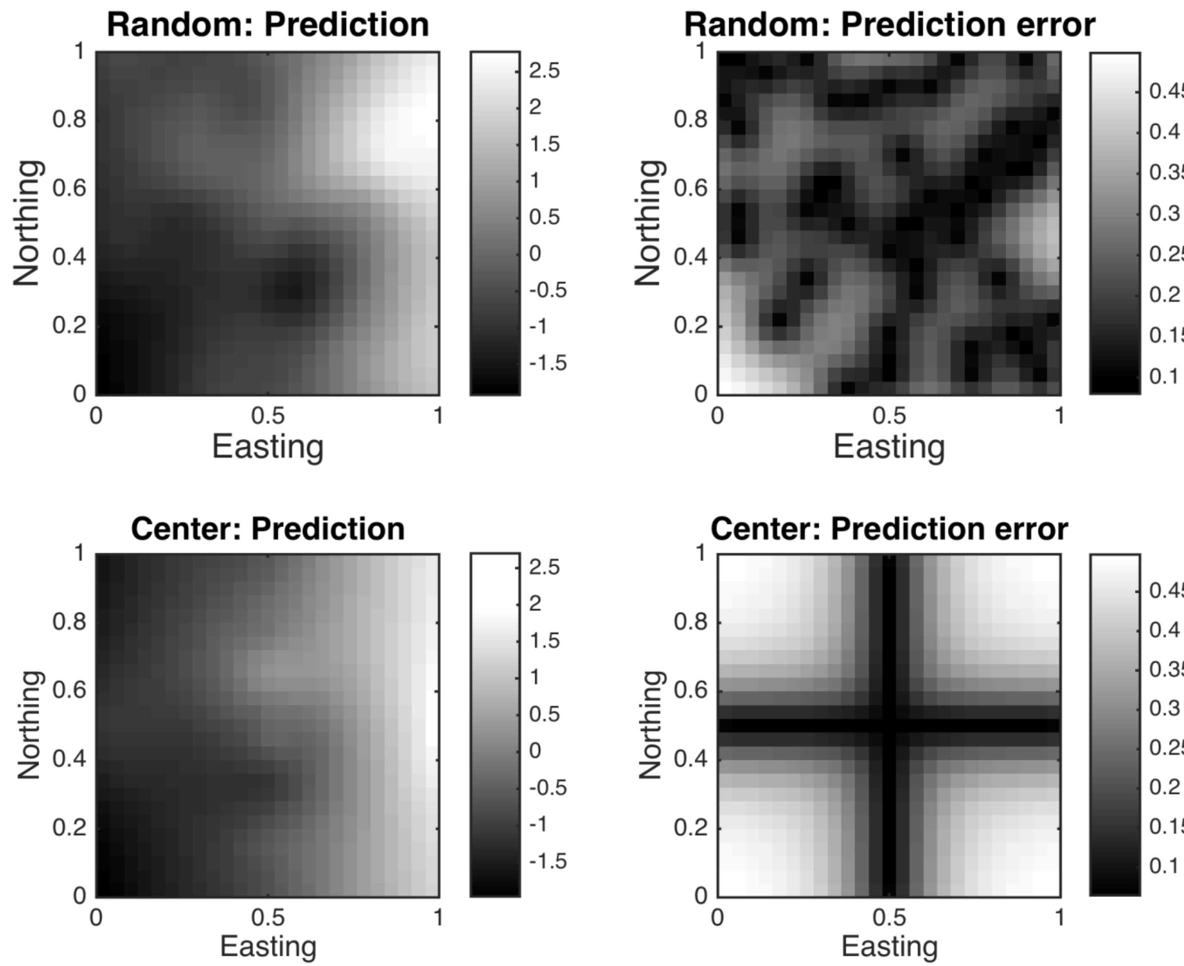
Goal is:

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

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

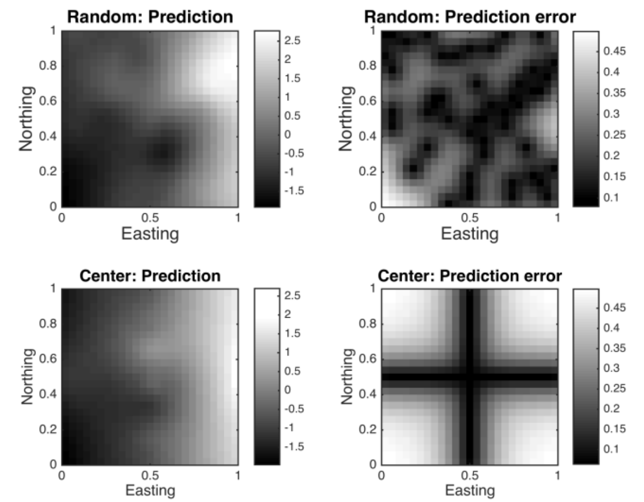
This is Kriging prediction and associated variance.

Norwegian wood – posterior results



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

Norwegian wood – information



Data gathering schemes / design / active learning
Can be based on different criteria :

- 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.
- Maximum variance reduction
- Maximum entropy
- Prediction error / Excursion sets
- Value of information (VOI)

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

Variance reduction

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

Expected variance reduction:

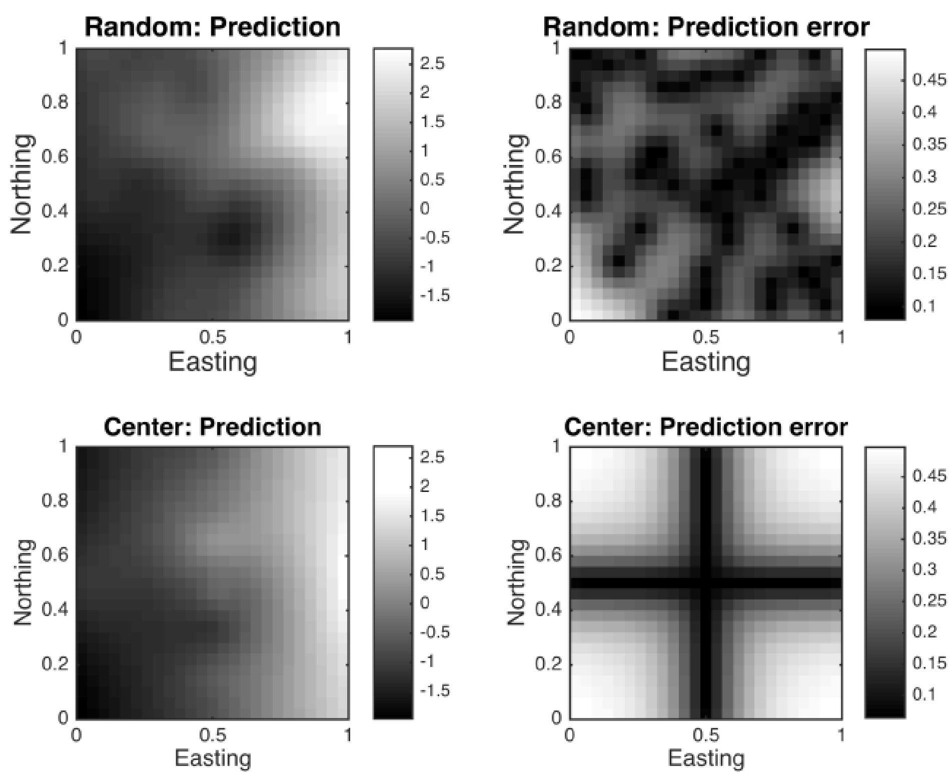
$$\text{EVR}(\boldsymbol{y}) = \sum_{i=1}^{n} Var(x_i) - E\left(\sum_{i=1}^{n} Var(x_i \mid \boldsymbol{y})\right) = \sum_{i=1}^{n} Var(x_i) - \int \sum_{i=1}^{n} Var(x_i \mid \boldsymbol{y}) p\left(\boldsymbol{y}\right) d\boldsymbol{y}$$

For the Gaussian distribution: $Var\left(\boldsymbol{x} \mid \boldsymbol{y}\right) = \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\boldsymbol{F}^t\left(\boldsymbol{F}\boldsymbol{\Sigma}\boldsymbol{F}^t + \tau^2\boldsymbol{I}\right)^{-1}\boldsymbol{F}\boldsymbol{\Sigma}$

The posterior variance does not depend on the data, only the data design.

Variance reduction (Kriging)

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



Overall variance reduction is larger for the random design.

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

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

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

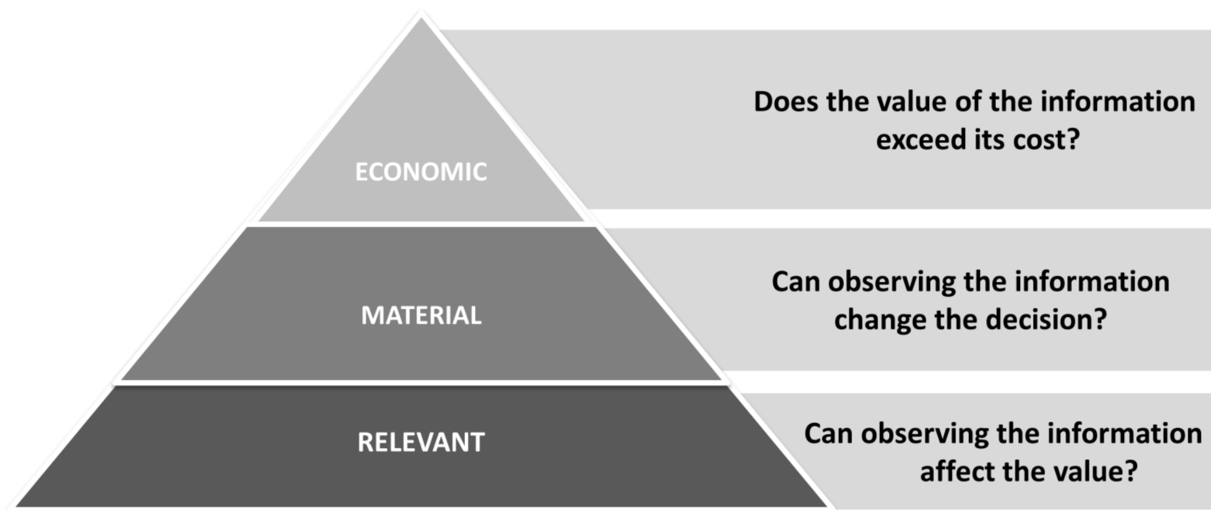
$$\boldsymbol{\Sigma} = 1.5\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

less entropy than

$$\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Since the entropy is the log determinant of the covariance matrix, the posterior entropy will not depend on the data, only the data design.

VOI - Pyramid of conditions



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

Formula for VOI

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

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

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

Decoupling – values are sums

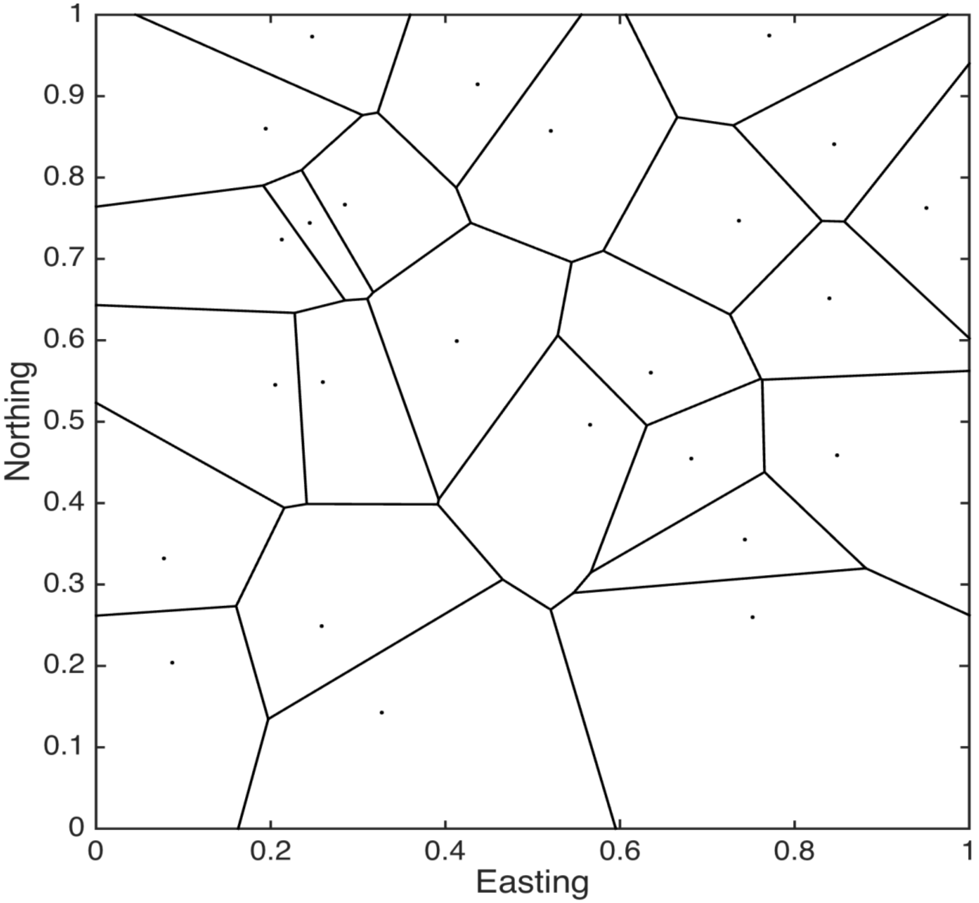
	Assumption: Decision Flexibility	**Assumption: Value Function**
Low decision flexibility; Decoupled value	Alternatives are easily enumerated $a \in A$	Total value is a sum of value at every unit $v(\mathbf{x},a)=\sum_j v(x_j,a)$
High decision flexibility; Decoupled value	None $\boldsymbol{a} \in \boldsymbol{A}$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

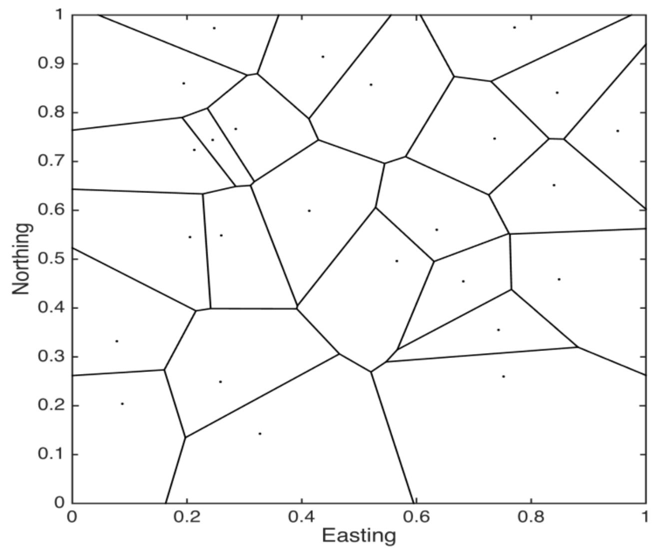
Low flexibility:
Farmer must select all forest units, or none.



Decoupled versus coupled value

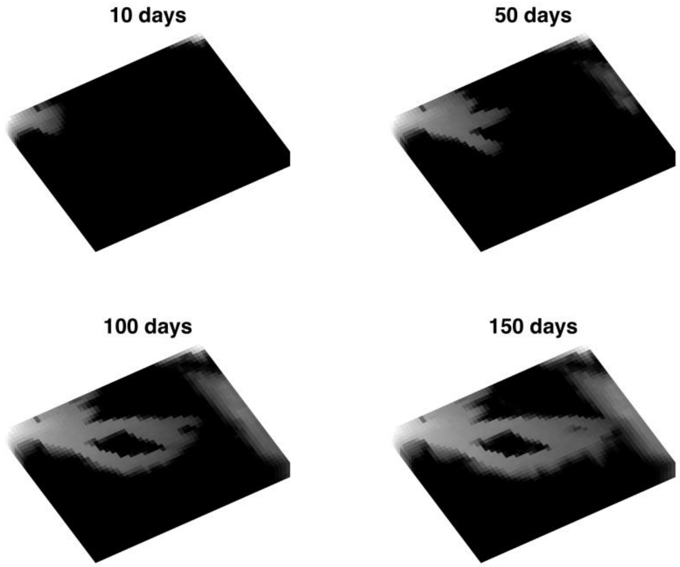
Farmer must decide whether to harvest at forest units, or not.

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

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

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

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 (**GAUSSIAN**).
- Otherwise approximations and Monte Carlo.

Conditioning – Gaussian models

Prediction, Kriging.

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

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

$$\boldsymbol{y} \sim N\left(\boldsymbol{F \mu}, \tau^2 \boldsymbol{I} + \boldsymbol{F \Sigma F}^t\right)$$

- Conditional mean is linear in the data.
- Data are Gaussian Linear combinations of Gaussian variables are Gaussian

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.
Value function is linear.
Profits modeled directly.

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

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

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

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

VOI – analytical expression

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

$$= \int \max\{0, w\}\, p(w)\, dw$$

$$w \sim N\left(\mu_w, r_w^2\right),$$

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

$$\boldsymbol{y} \sim N\left(\boldsymbol{F\mu}, \tau^2 \boldsymbol{I} + \boldsymbol{F \Sigma F}^t\right)$$

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

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

VOI – analytical expression

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

$$= \int \max\{0, w\}\, p(w)\, dw = \mu_w \Phi\left({\mu_w}/{r_w} \right) + r_w \phi\left({\mu_w}/{r_w} \right)$$

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

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

$$= \mu_w \Phi\left({\mu_w}/{r_w} \right) + r_w \phi\left({\mu_w}/{r_w} \right) - \max\left\{0, \sum_{i=1}^{n} \mu_i \right\}$$

Design

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

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

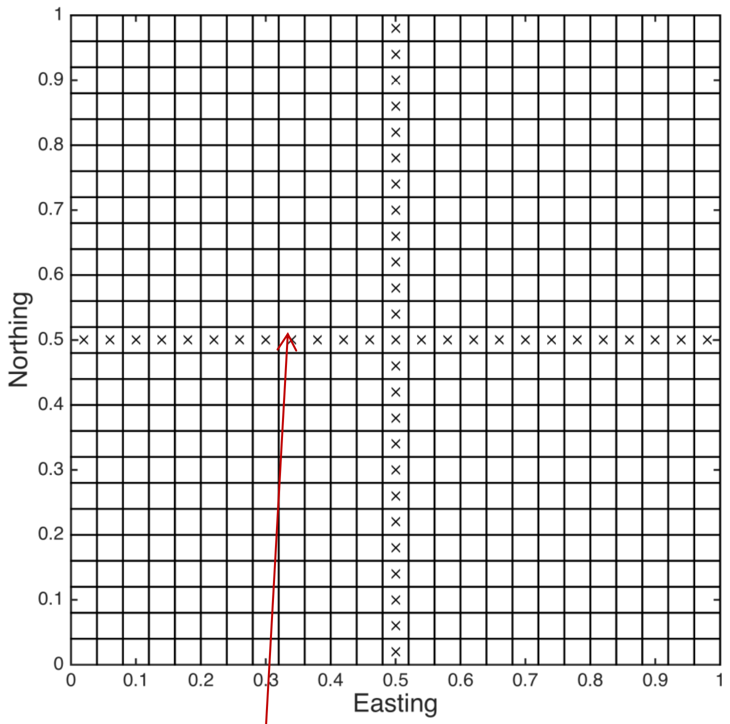
$$\boldsymbol{F} = \boldsymbol{I}$$

Total test.

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

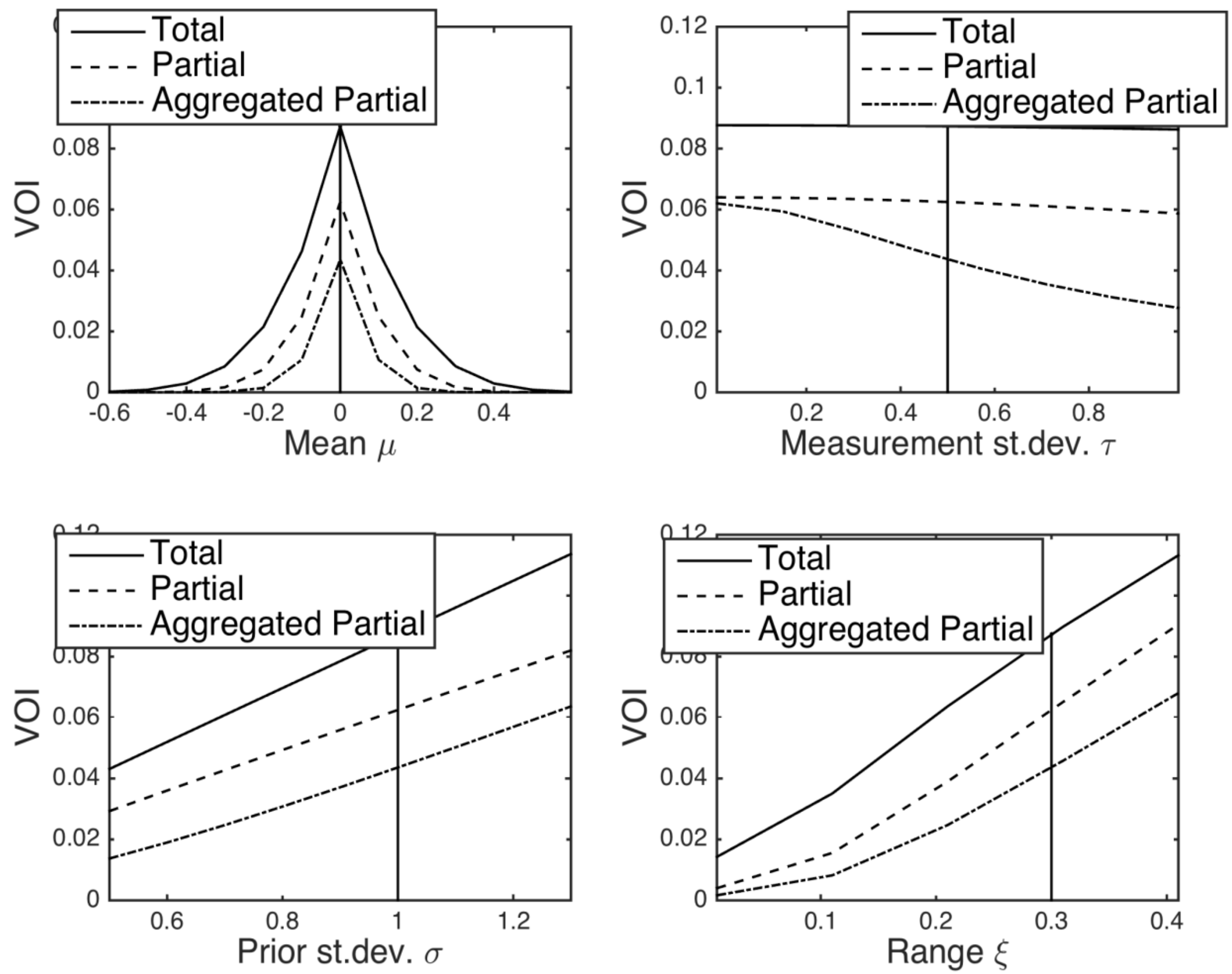
$$p\left(\boldsymbol{y} \mid \boldsymbol{x}\right) = N\left(\boldsymbol{F}\boldsymbol{x}, \tau^2 \boldsymbol{I}\right)$$

Example of imperfect test.



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

Results - Forestry example

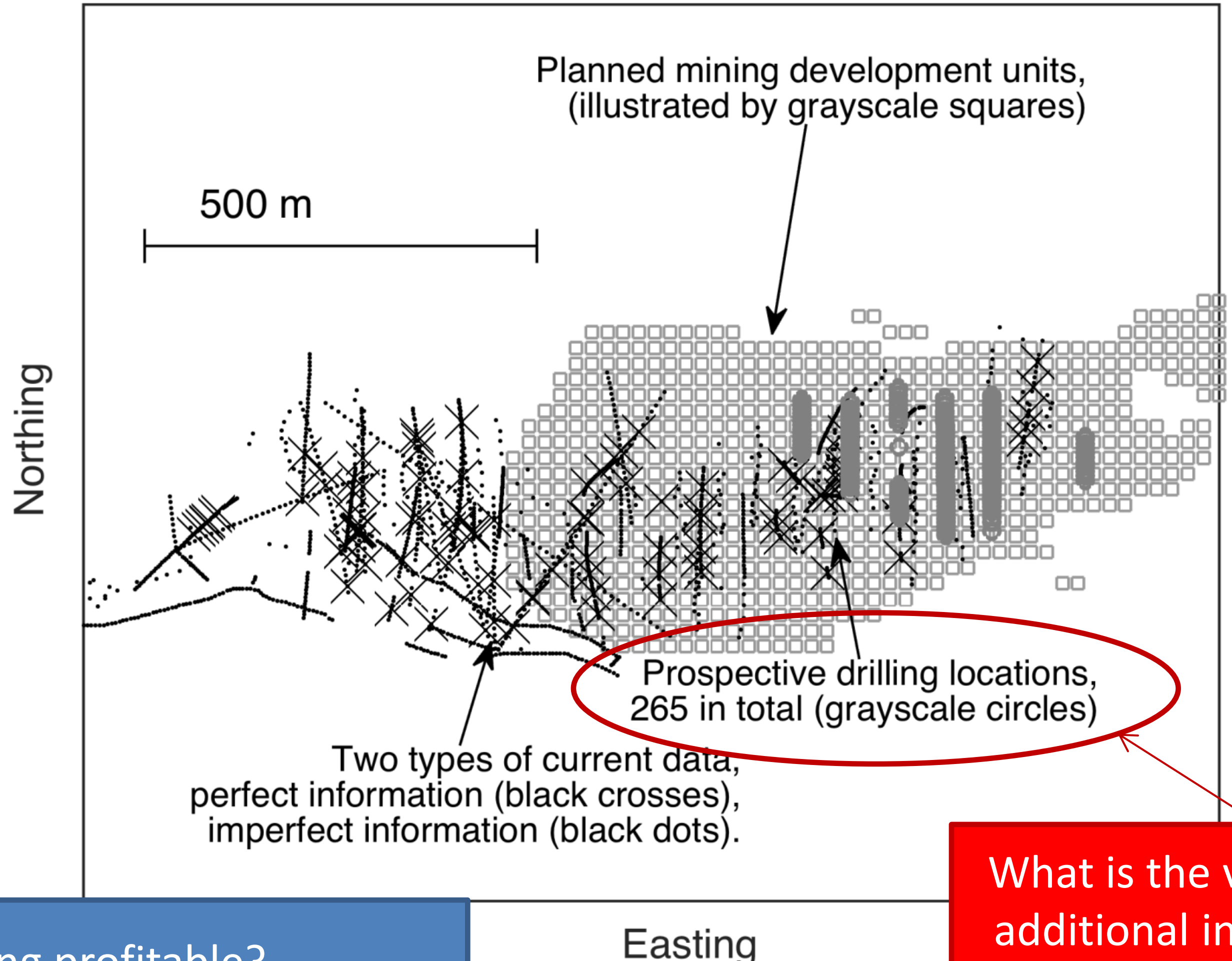


Total: all cells. Partial: Every cell along center lines. Aggregated partial: sums along center lines. (Results are normalized for area).

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.

I love rock and ore – mining example

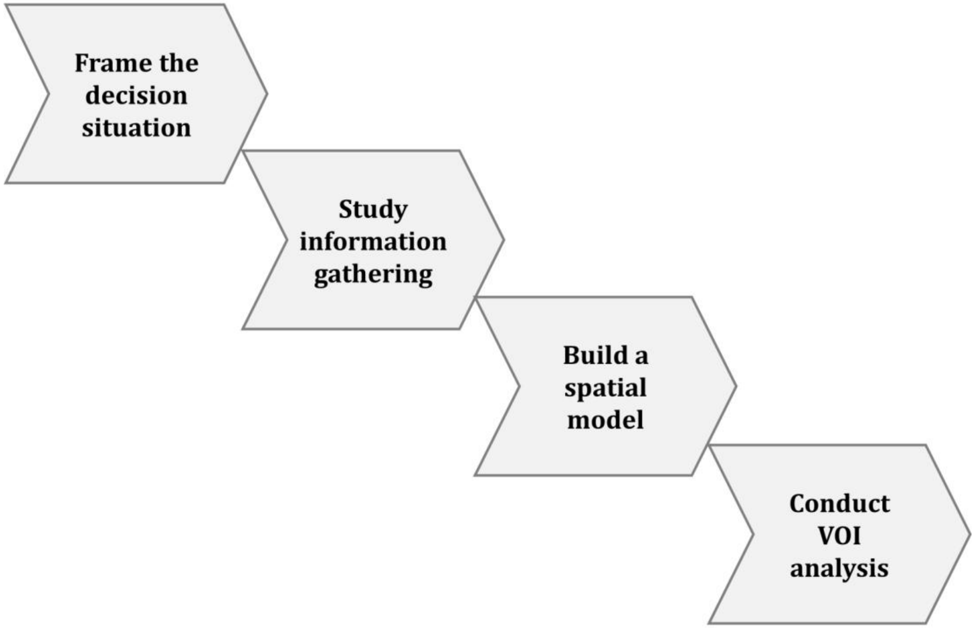


What is the value of this additional information?

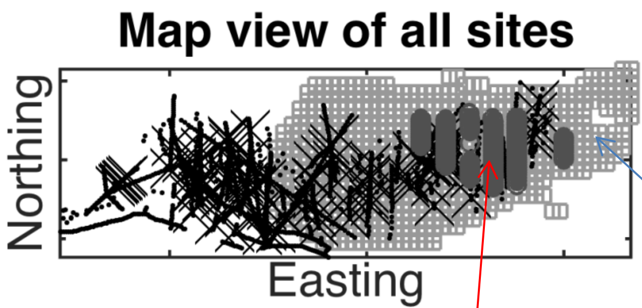
Is mining profitable?

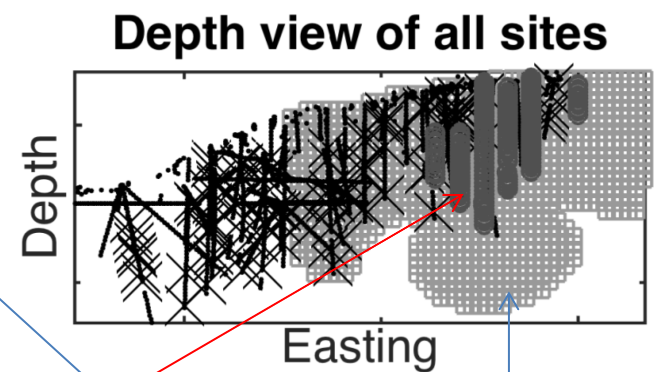
VOI workflow

- Low decision flexibility. De-coupled value function. Some of all units.
- 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

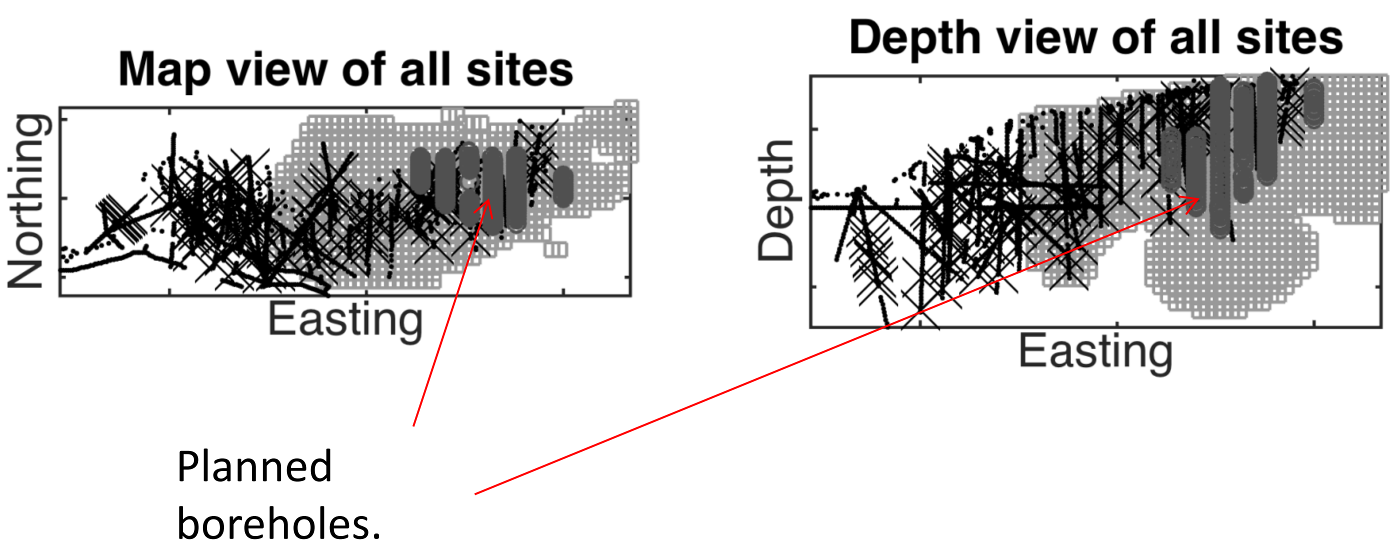




Planned boreholes.

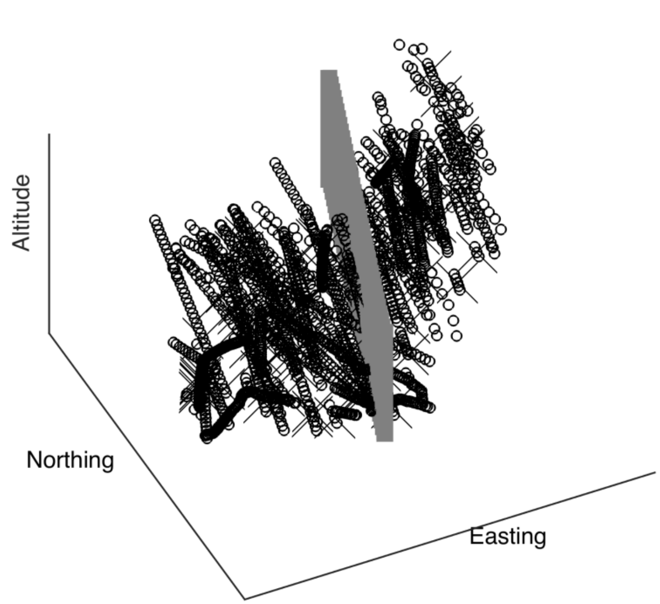
Mining blocks. Some waste rock. Some high-grade.

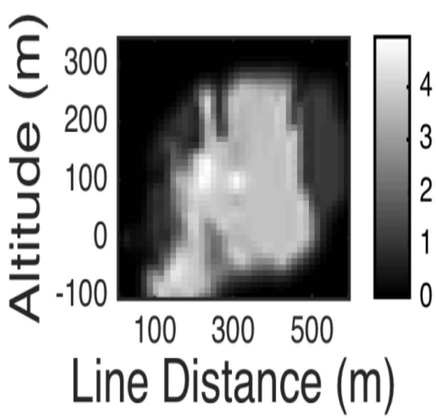
Information gathering

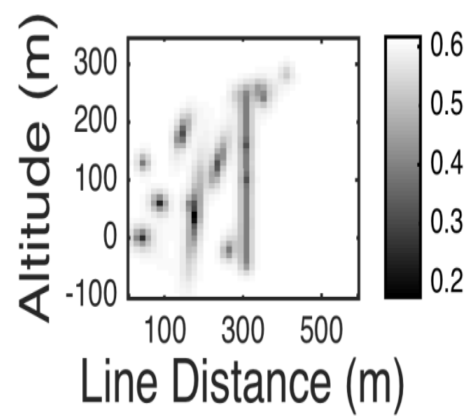


- **Total test** : 265 measurements in 21 new boreholes.
- **Partial test**: Drilling and sampling data only in a subset of boreholes.
- **Perfect** testing (XRF: done in lab). **Imperfect** testing (XMET: handheld meter).

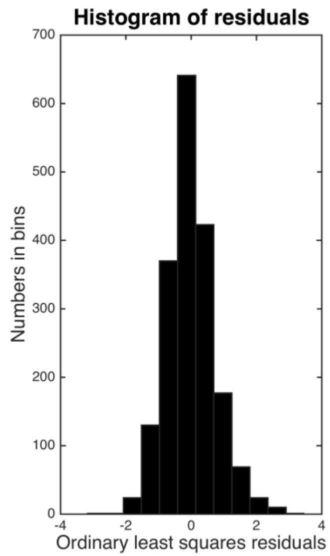
Prior model

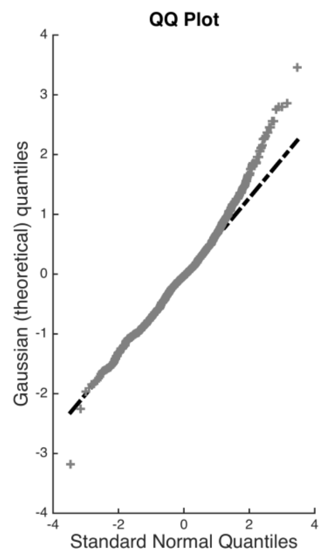

XMET
XRF
Altitude
Northing
Easting

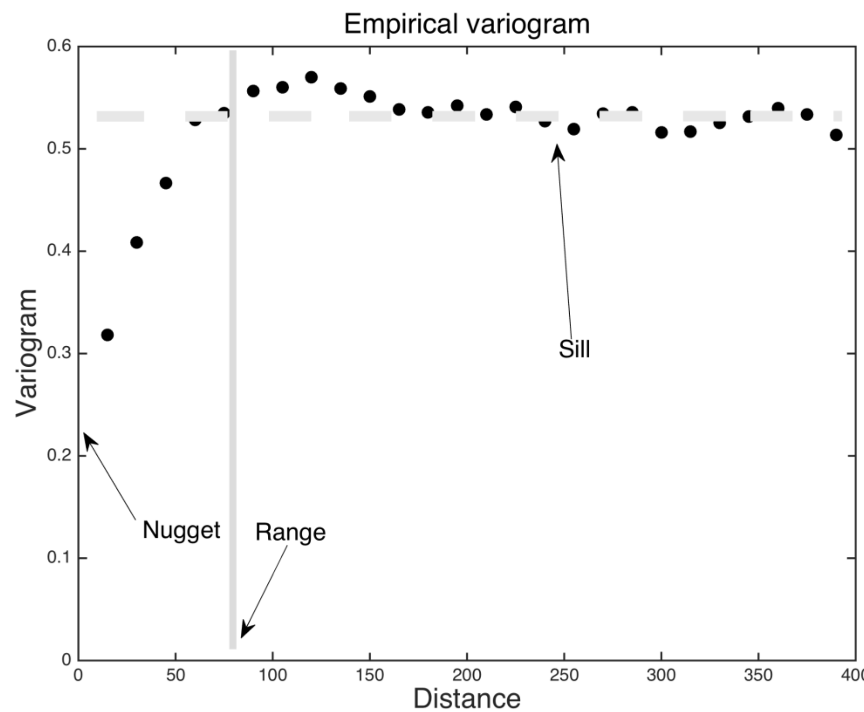

Altitude (m)
300
200
100
0
-100
100
300
500
Line Distance (m)
4
3
2
1
0


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
0
100
200
300
400
500
600
700
-4
-2
0
2
4


QQ Plot
Gaussian (theoretical) quantiles
Standard Normal Quantiles
-4
-3
-2
-1
0
1
2
3
4


Empirical variogram
Variogram
Distance
Sill
Nugget
Range
0
0.1
0.2
0.3
0.4
0.5
0.6
50
100
150
200
250
300
350
400

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

VOI

Weights set from block model(waste or ore).

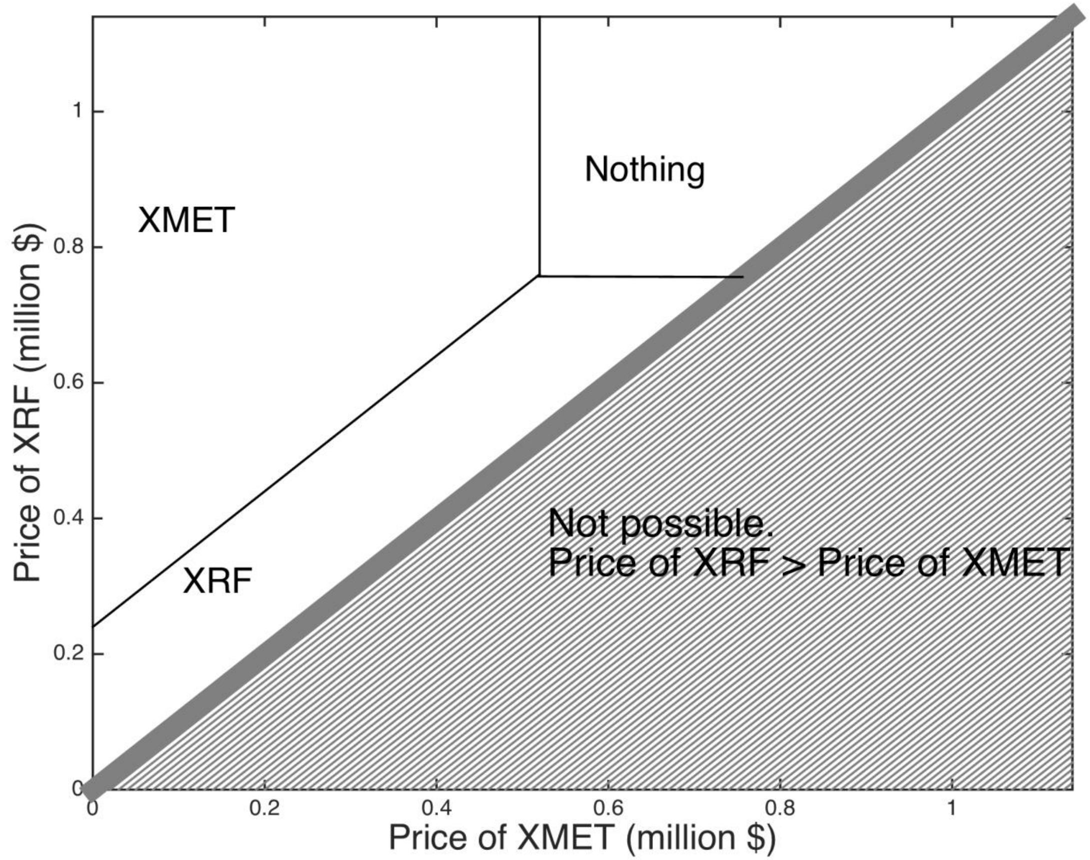
$$PV = \max\left\{E\left(\boldsymbol{w}^t\boldsymbol{x} - \text{Cost}\right), 0\right\}$$

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

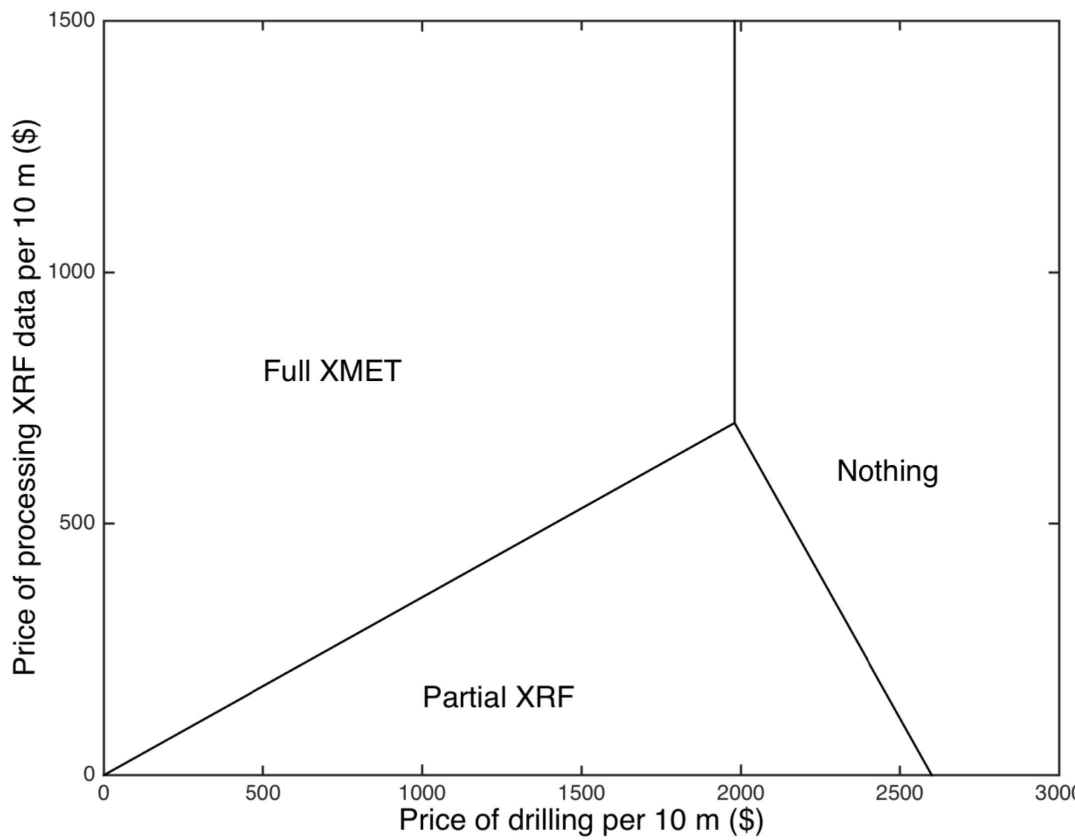
$$VOI(\boldsymbol{y}) = PoV(\boldsymbol{y}) - PV.$$

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).