Day 3: Decision analyses

Introduction and fundamentals of Bayesian decision analysis

- Prior, Pre-posterior and Posterior decision analysis
- Quantification of utilities

Value of Information analyses and decision analysis types

- Types of Value of Information analyses
- Decision analyses types
- Derivation of decision rules

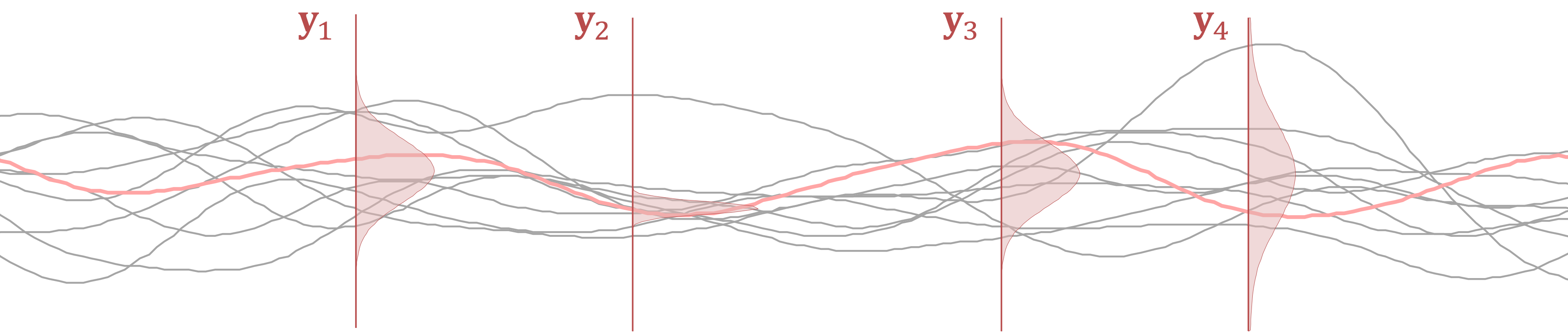
Value of Information analyses cont.

- Introduction to Gaussian random field models
- Evaluation of Value of Information in Gaussian random fields
- Optimal sensors placement and inspection scheduling using Value of Information

Carl Malings, Matteo Pozzi

decision analysis types and value of information analyses – part 2

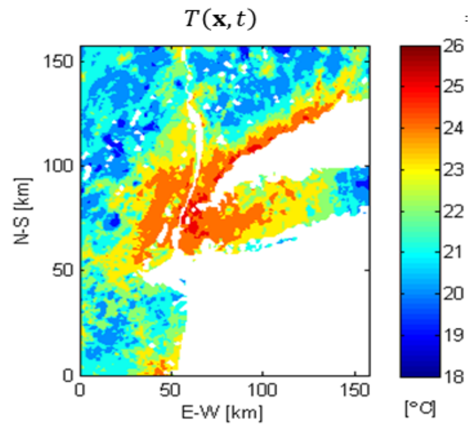
VoI in spatially distributed systems



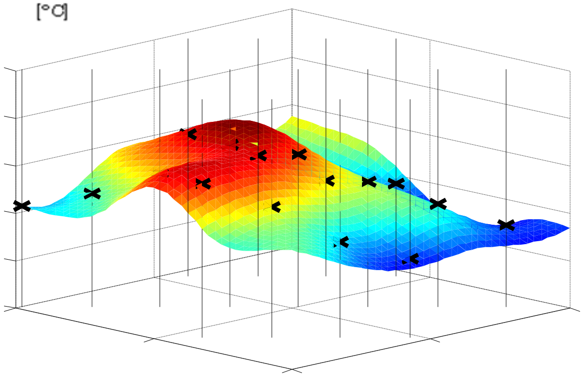
Outline

- **Random Field Models**
  - Conditional probability and Bayesian updating
  - Introduction to Gaussian random field models
- **Value of Information in Random Field Models**
  - Efficient evaluation of value of information in Gaussian Random fields
  - Sensor placement optimization using value of information
- **Extension to Spatio-Temporal Systems**
  - Value of Information for sequential decision-making
  - Optimal placement and scheduling of measurements

Gaussian Fields for Spatially Distributed Phenomena



- seismic demand
- temperature prediction
- corrosion of concrete slabs
- permeability in soil
- …
- *any spatiallly distributed phenomena*
- …



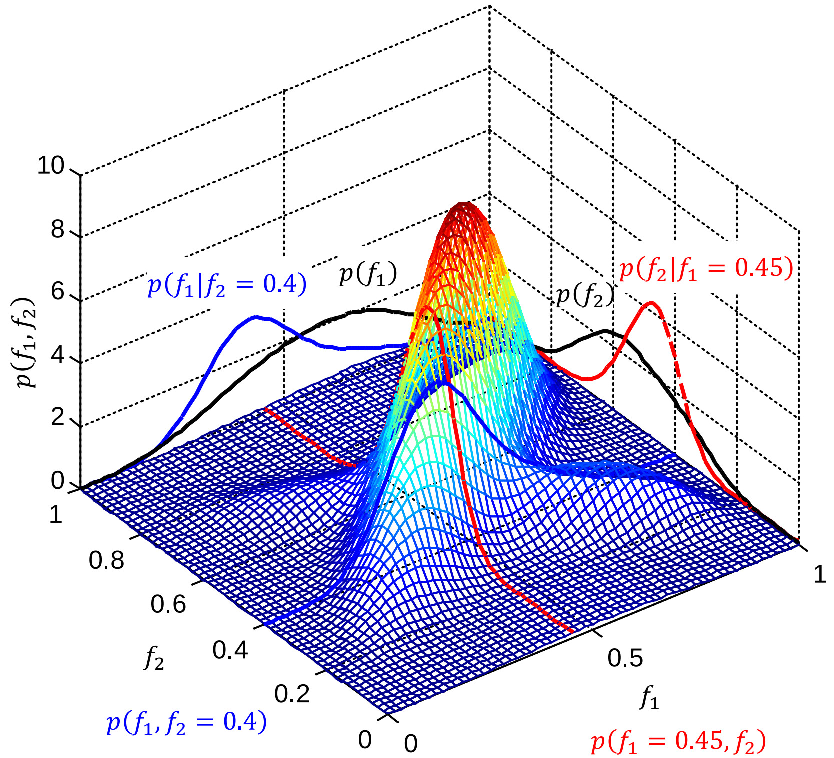
- supervise learning in Machine Learning, non-linear fitting
- related to stochastic processes in time and space domain, e.g. in random vibration analysis

Distributions for Continuous Variables

joint distribution:

$p(f_1, f_2)$ $\qquad$ $p(f_1, f_2) \geq 0$ *positive*

normalized $\displaystyle\iint_{-\infty}^{\infty} p(f_1, f_2) df_1 df_2 = 1$



marginal distributions:

$$p(f_1) = \int_{-\infty}^{\infty} p(f_1, f_2) df_2$$

$$p(f_2) = \int_{-\infty}^{\infty} p(f_1, f_2) df_1$$

conditional:

$$p(f_1|f_2) = \frac{p(f_1, f_2)}{p(f_2)}$$

$$p(f_2|f_1) = \frac{p(f_1, f_2)}{p(f_1)}$$

Bayesian Inference for Continuous Variables

model the "whole world"

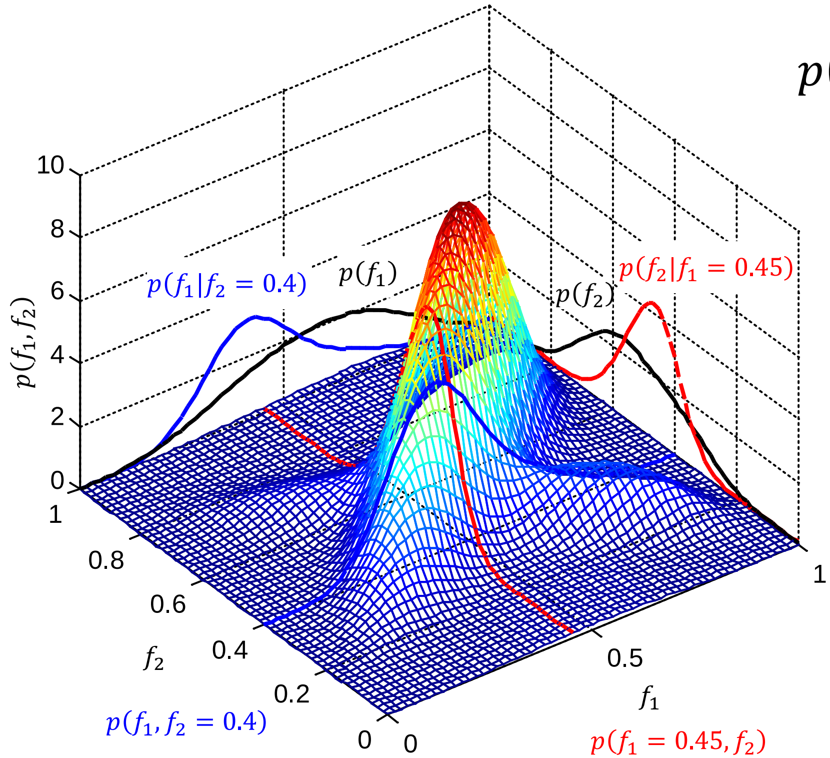
$p(f_1, f_2)$

observe something

$f_2 = \tilde{f}$

take a section

$$\begin{cases} f_2 = \tilde{f} & \text{not a rand. var. anymore} \\ p(f_1 | f_2 = \tilde{f}\,) \end{cases}$$



$$p(f_1 | f_2 = \tilde{f}) = \frac{p(f_1, f_2 = \tilde{f})}{p(f_2 = \tilde{f})}$$

$$p(f_2 = \tilde{f}) = \int_{-\infty}^{\infty} p(f_1, f_2 = \tilde{f}) df_1$$

inference is relevant only for dependent variables

$$\exists f_2 \colon p(f_1 | f_2) \neq p(f_1)$$

Bayesian Inference for Independent Variables

model the "whole world"

$p(f_1, f_2)$

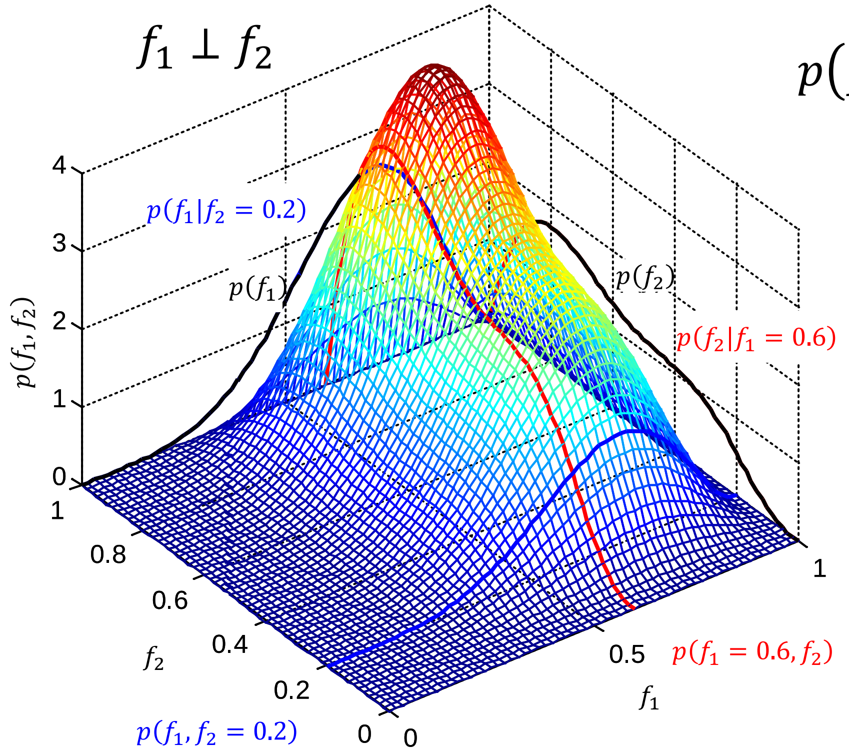
observe something

$f_2 = \tilde{f}$

take a section

$$\begin{cases} f_2 = \tilde{f} & \text{not a rand. var. anymore} \\ p(f_1 | f_2 = \tilde{f}\,) \end{cases}$$

$f_1 \perp f_2$



$$p(f_1 | f_2 = \tilde{f}) = \frac{p(f_1, f_2 = \tilde{f})}{p(f_2 = \tilde{f})}$$

$$p(f_2 = \tilde{f}) = \int_{-\infty}^{\infty} p(f_1, f_2 = \tilde{f})\, df_1$$

$f_1 \perp f_2$

$p(f_1|f_2) = p(f_1)$

no updating

$$\begin{aligned} p(f_1, f_2) &= p(f_2|f_1)p(f_1) \\ &= p(f_2)p(f_1) \end{aligned}$$

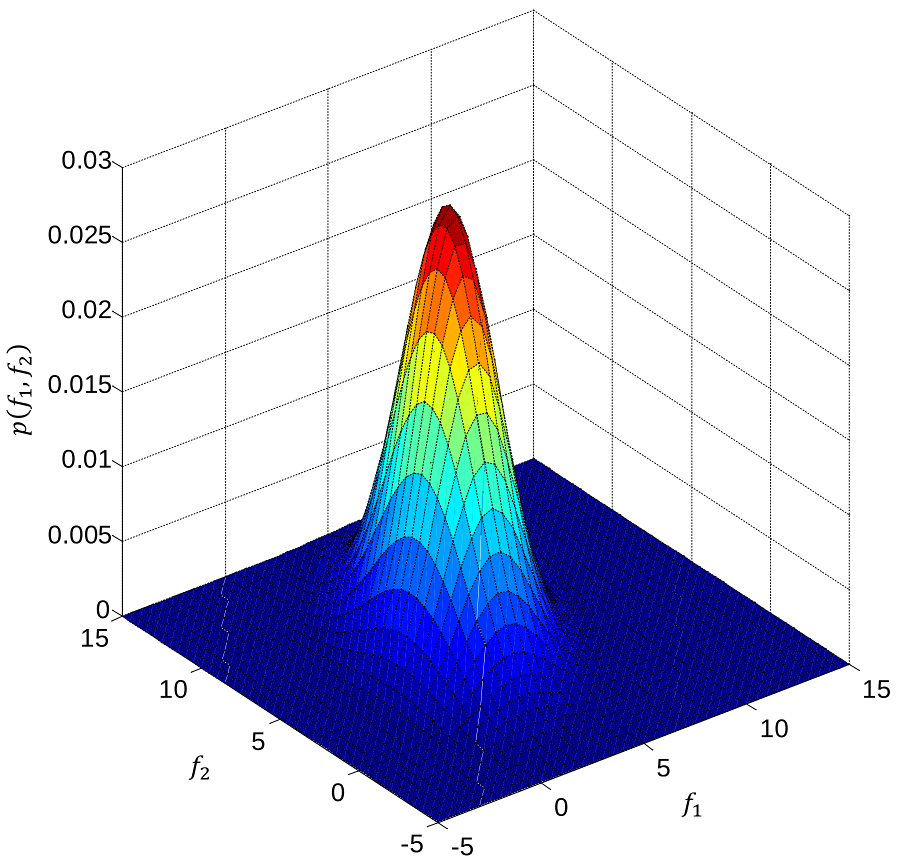
Multivariate Normal Distribution

pdf: $$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{|\boldsymbol{\Sigma}|}} \cdot \exp\left[-\frac{1}{2}(\mathbf{f}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{f}-\boldsymbol{\mu})\right] \qquad \mathbb{R}^n \longrightarrow \mathbb{R}^+$$

vector of random variables | mean vector | covariance matrix

- the joint probability is completely defined by mean vector and covariance matrix, which are the parameters of the distribution.
- the conditional distribution, given any subset of variables, is also jointly normal.
- each subset of $\mathbf{f}$ is jointly normally distributed, and marginalization is computationally trivial (just copy part of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).
- any linear transformation of the variables is jointly normal.

$\mu_1 = 4 \quad \mu_2 = 6 \quad \sigma_1 = 2 \quad \sigma_2 = 3 \quad \rho_{12} = 0.4$



Normal Model

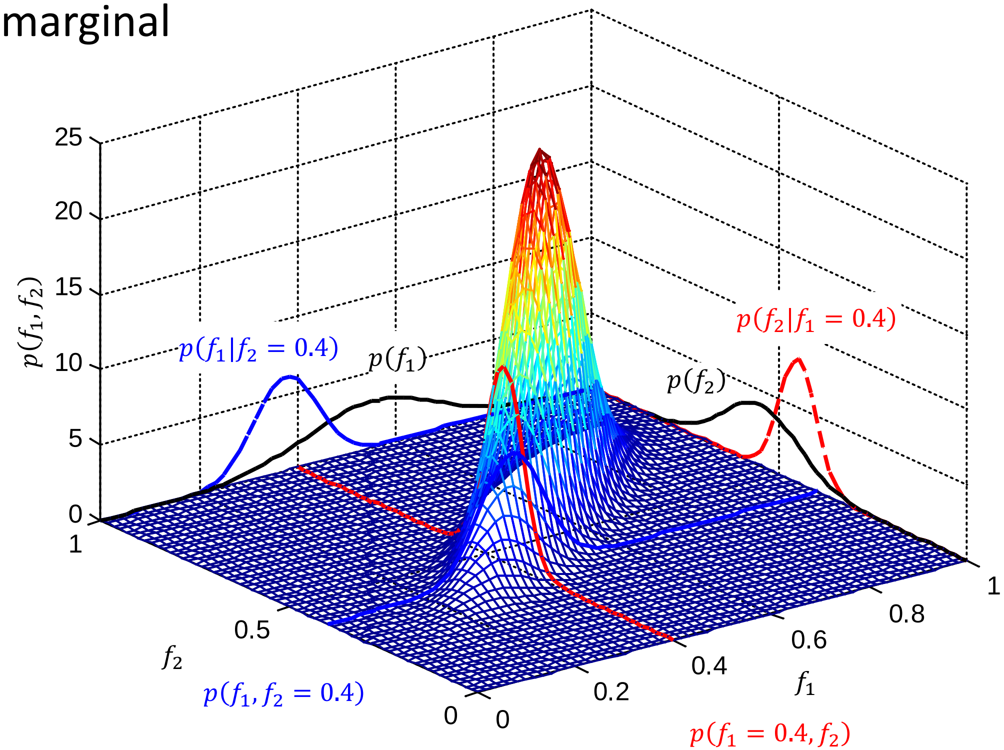
joint

$$p(f_1, f_2) = \mathcal{N} \longrightarrow p(f_1, f_2 = \tilde{f}) \propto \mathcal{N} \longrightarrow p(f_1 | f_2 = \tilde{f}) = \frac{p(f_1, f_2 = \tilde{f})}{p(f = \tilde{f})} = \mathcal{N}$$

conditional

$$p(f_2 = \tilde{f}) = \int_{-\infty}^{\infty} p(f_1, f_2 = \tilde{f}) df_1 \propto \mathcal{N}$$

marginal



Normal joint density

→ Normal marginal density

→ Normal conditional density

Bayesian Updating in Normal Models

perfect observations

prior prob.

$$\begin{bmatrix}\mathbf{f}_1\\ \mathbf{f}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_{F_1}\\ \boldsymbol{\mu}_{F_2}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_{F_1,F_1} & \boldsymbol{\Sigma}_{F_1,F_2}\\ \boldsymbol{\Sigma}_{F_2,F_1} & \boldsymbol{\Sigma}_{F_2,F_2}\end{bmatrix}\right)$$

conditional after observing $\mathbf{f}_2$

$$\mathbf{f}_1|\mathbf{f}_2 \sim \mathcal{N}\left(\boldsymbol{\mu}_{F_1|\mathbf{f}_2}, \boldsymbol{\Sigma}_{F_1|F_2}\right)$$

$$\boldsymbol{\mu}_{F_1|\mathbf{f}_2} = \boldsymbol{\mu}_{F_1} + \boldsymbol{\Sigma}_{F_1,F_2}\boldsymbol{\Sigma}_{F_2,F_2}{}^{-1}(\mathbf{f}_2 - \boldsymbol{\mu}_{F_2})$$

$$\boldsymbol{\Sigma}_{F_1|F_2} = \boldsymbol{\Sigma}_{F_1,F_1} - \boldsymbol{\Sigma}_{F_1,F_2}\boldsymbol{\Sigma}_{F_2,F_2}{}^{-1}\boldsymbol{\Sigma}_{F_2,F_1}$$

imperfect observations

prior prob. linear observation

$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}_F, \boldsymbol{\Sigma}_F) \qquad \mathbf{y} = \mathbf{R}\mathbf{f} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}_\epsilon, \boldsymbol{\Sigma}_\epsilon)$$

joint distribution

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{y}\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_F\\ \boxed{\mathbf{R}\boldsymbol{\mu}_F + \boldsymbol{\mu}_\epsilon}\end{bmatrix}, \begin{bmatrix}\boldsymbol{\Sigma}_F & \boldsymbol{\Sigma}_F\mathbf{R}^{\mathrm{T}}\\ \mathbf{R}\boldsymbol{\Sigma}_F & \boxed{\mathbf{R}\boldsymbol{\Sigma}_F\mathbf{R}^{\mathrm{T}} + \boldsymbol{\Sigma}_\epsilon}\end{bmatrix}\right)$$

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$$

posterior $\mathbf{f}|\mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}_{F|\mathbf{y}}, \boldsymbol{\Sigma}_{F|Y}\right)$

$$\boldsymbol{\mu}_{F|\mathbf{y}} = \boldsymbol{\mu}_F + \boldsymbol{\Sigma}_F\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}_Y{}^{-1}(\mathbf{y} - \boldsymbol{\mu}_Y)$$

$$\boldsymbol{\Sigma}_{F|Y} = \boldsymbol{\Sigma}_F - \boldsymbol{\Sigma}_F\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}_Y{}^{-1}\mathbf{R}\boldsymbol{\Sigma}_F$$

e.g. y_1 measures $f_1 + f_3$ with noise σ_{ϵ_1}
y_2 measures ${}^1/_3\, f_2 + 5$ with noise σ_{ϵ_2}

$$\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}1 & 0 & 1\\ 0 & {}^1/_3 & 0\end{bmatrix}\begin{bmatrix}f_1\\ f_2\\ f_3\end{bmatrix} + \begin{bmatrix}\boldsymbol{\epsilon}_1\\ \boldsymbol{\epsilon}_2\end{bmatrix}$$

$$\begin{bmatrix}\boldsymbol{\epsilon}_1\\ \boldsymbol{\epsilon}_2\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\ 5\end{bmatrix}, \begin{bmatrix}\sigma_{\epsilon_1}^2 & 0\\ 0 & \sigma_{\epsilon_2}^2\end{bmatrix}\right)$$

for normal vars. inference can be performed in closed-form

Temperature in Two Rooms

room 1

$f_1 = f(x_1)$

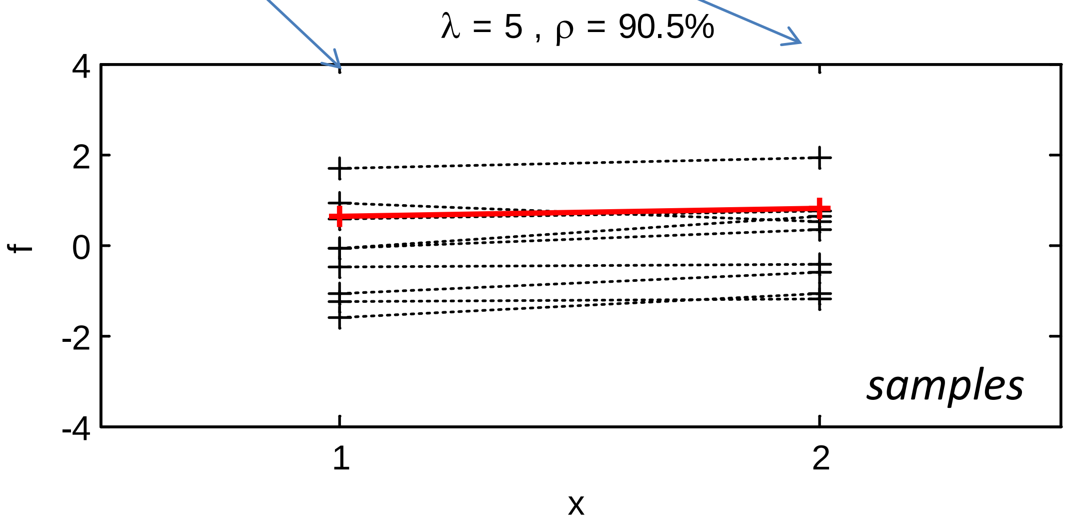
room 2

$f_2 = f(x_2)$

same marginal: $i = 1,2: \quad f_i \sim \mathcal{N}\left(0, \sigma_f^2\right)$

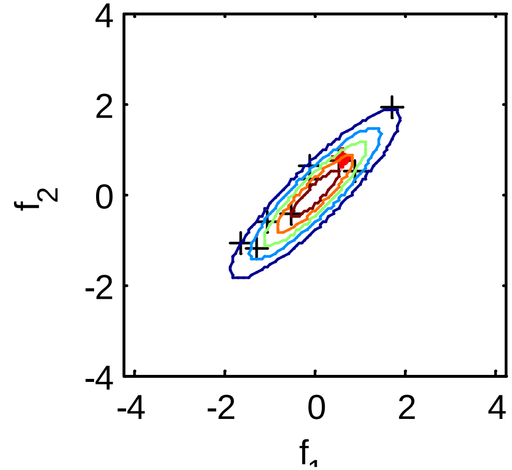
$\mu_i = 0\ ; \quad \sigma_i = 1$

samples:



joint probability

$p(f_1, f_2)$



Temperature in Three Rooms

room 1	room 2	room 3
$f_1 = f(x_1)$	$f_2 = f(x_2)$	$f_3 = f(x_3)$

ρ_{12} ρ_{23}

ρ_{13}

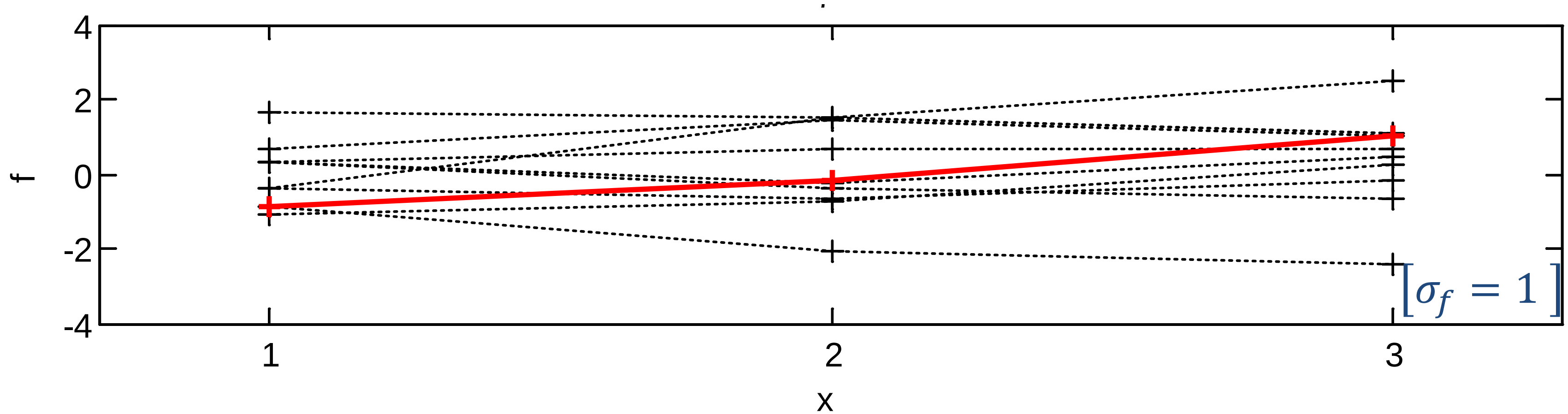
same marginal: $i = 1,2,3\text{:}\ f_i \sim \mathcal{N}\left(0, \sigma_f^2\right)$

$\mathbf{R}$: correlation matrix

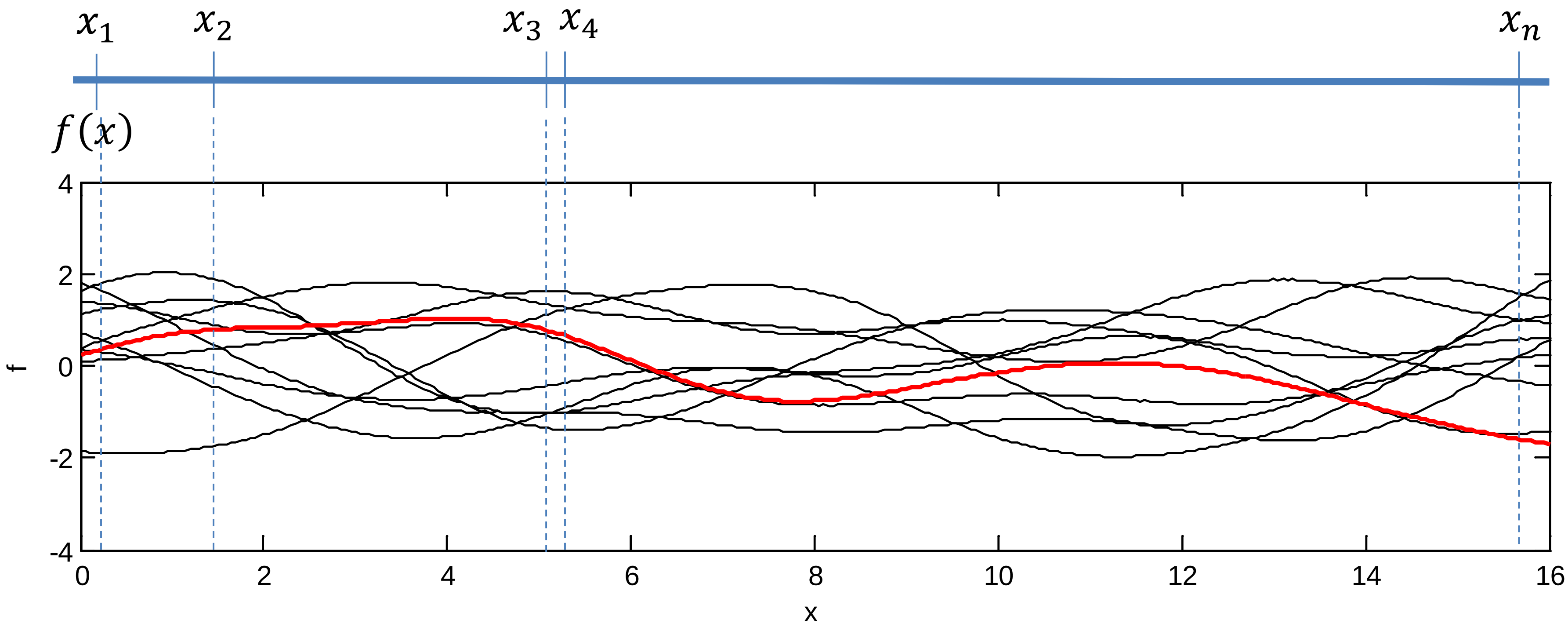
$$\mathbf{f} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}) \qquad \mathbf{\Sigma} = \sigma_f^2 \underbrace{\begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}}_{\mathbf{R}}$$

ρ_{12}, ρ_{23}: high because temperature in rooms places side-by-side is similar, e.g. 75%

ρ_{13} : lower. e.g. 33%



Temperature Along a Hall: “Continuous” Case



we can plot a function that looks continuous by including a thin grid in the analysis.

Again, if we are only interested in n points, all others are irrelevant: we can include them or not in our analysis.

Covariance Function: squared exponential

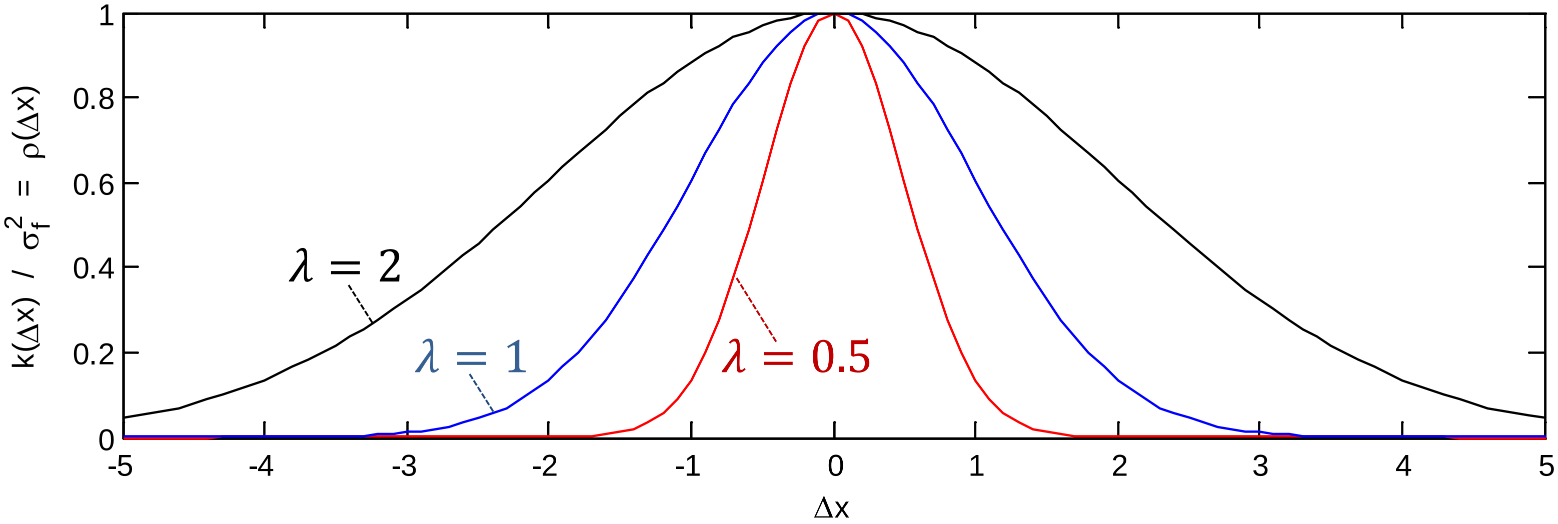
$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \mathbf{0}, \mathbf{\Sigma}) \qquad \mathbf{\Sigma} = \mathbf{K}(\mathbf{x}, \mathbf{x})$$

$$\Delta x = x_i - x_j$$

$$\sigma_{ij}^2 = \sigma_f^2 \rho_{ij} = k(x_i, x_j) = k(\|x_i - x_j\|) = \sigma_f^2 \exp\left[-\frac{1}{2\lambda^2}(x_i - x_j)^2\right]$$

λ is the length-scale

$$\rho(\lambda) = e^{-1/2} = 60.6\%$$



σ_{ij}^2 is always positive for this model, it decays with $\left(\frac{\Delta x}{\lambda}\right)^2$.

Covariance Function: squared exponential

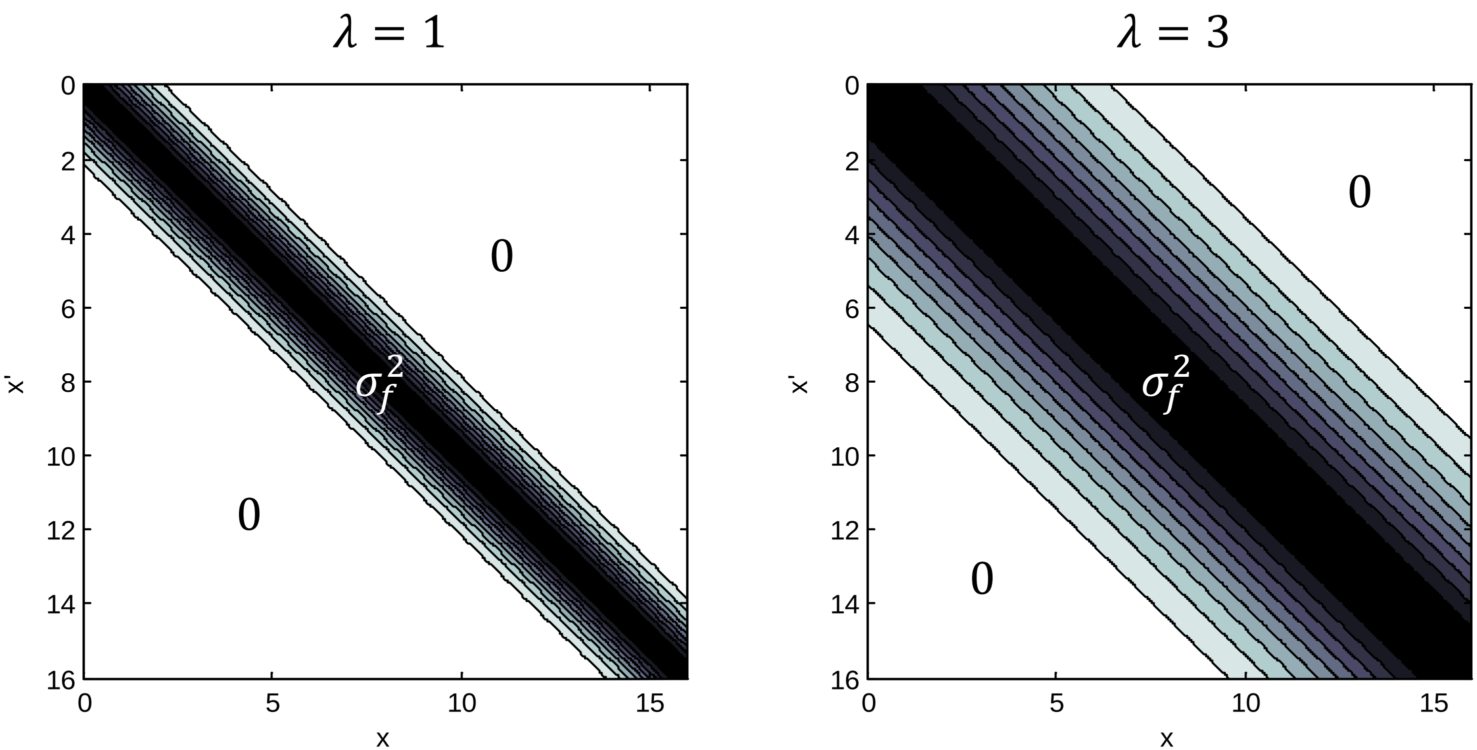
$$p(\mathbf{f}) = \mathcal{N}(\mathbf{f}; \mathbf{0}, \boldsymbol{\Sigma}) \qquad \boldsymbol{\Sigma} = \mathbf{K}(\mathbf{x}, \mathbf{x})$$

$\Delta x = x_i - x_j$

$$\sigma_{ij}^2 = \sigma_f^2 \rho_{ij} = k(x_i, x_j) = k(\|x_i - x_j\|) = \sigma_f^2 \exp\left[-\frac{1}{2\lambda^2}(x_i - x_j)^2\right]$$

λ is the length-scale

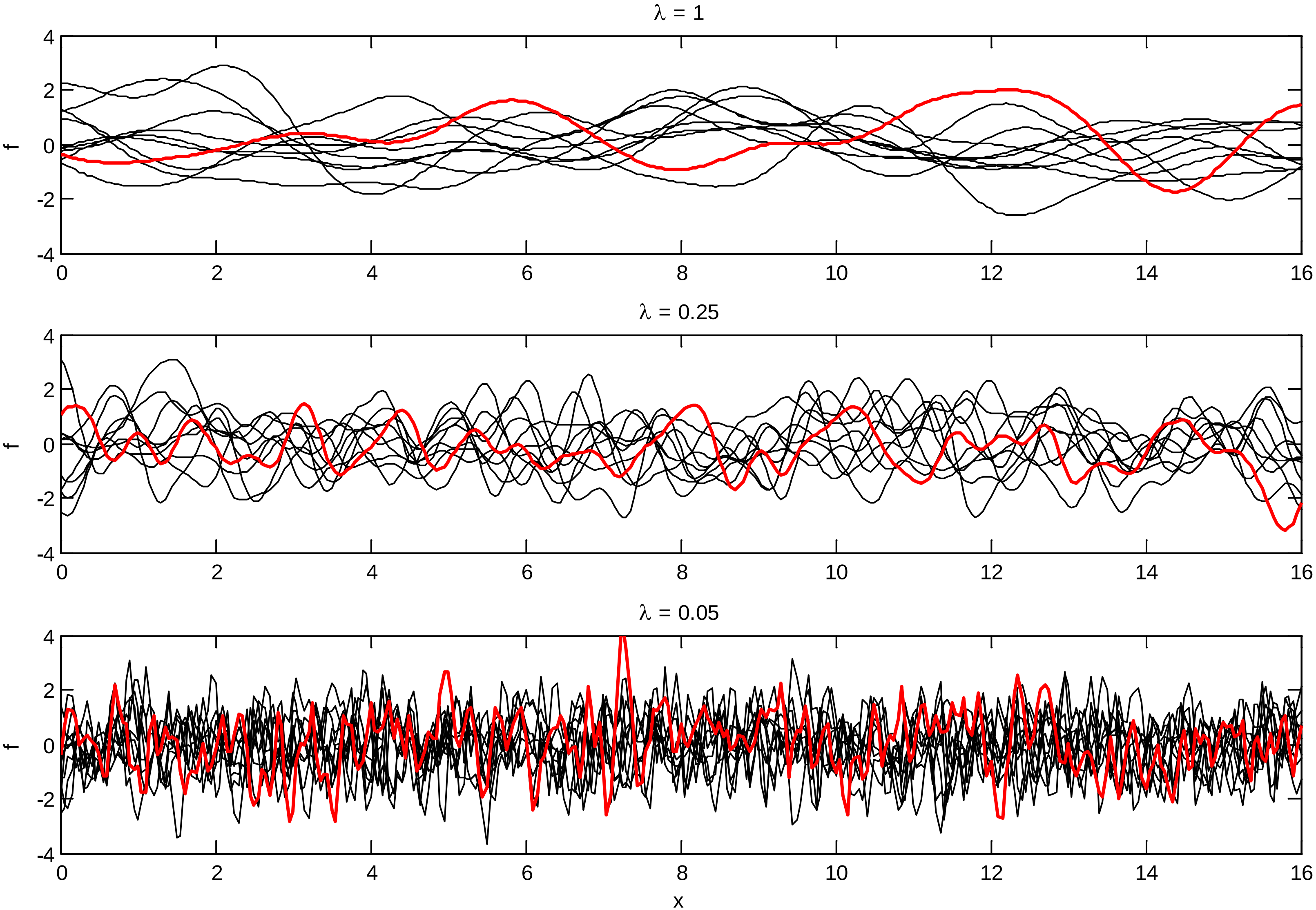
$k(x, x')$



The Role of Correlation Length

$\lambda \to \infty$

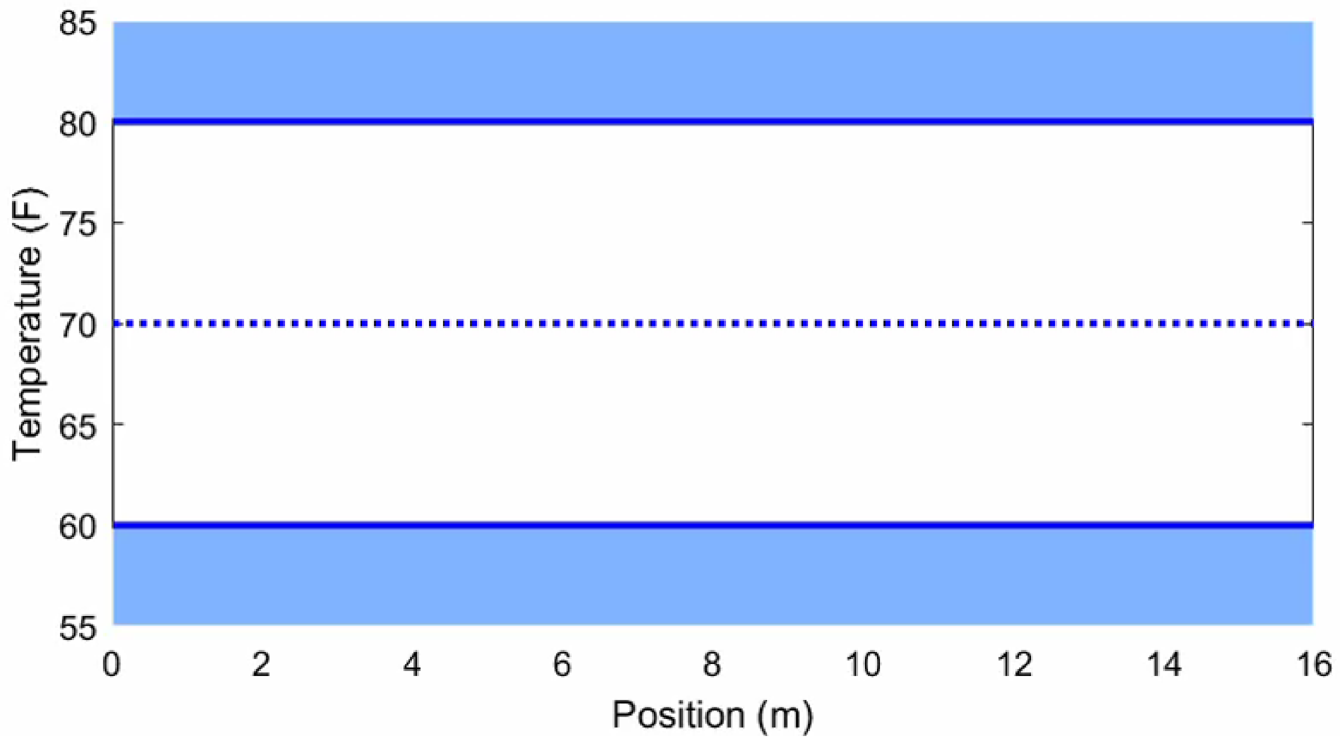
constant value



$\lambda \cong 0$

white noise

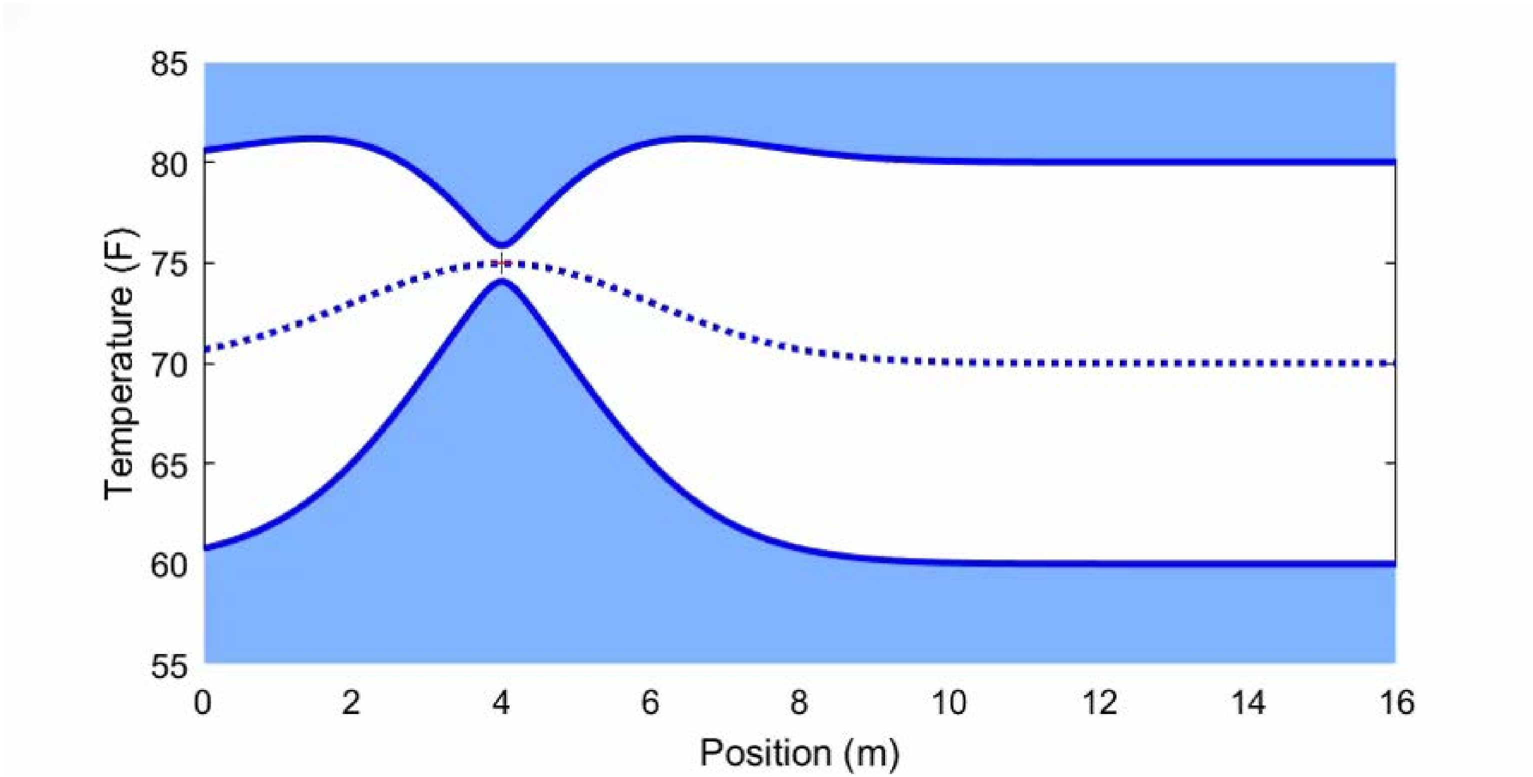
Spatially Distributed Phenomena



example: ambient temperature field

Gaussian process: $\mathcal{GP}\big(m(x); k(x, x')\big)$

Spatially Distributed Phenomena



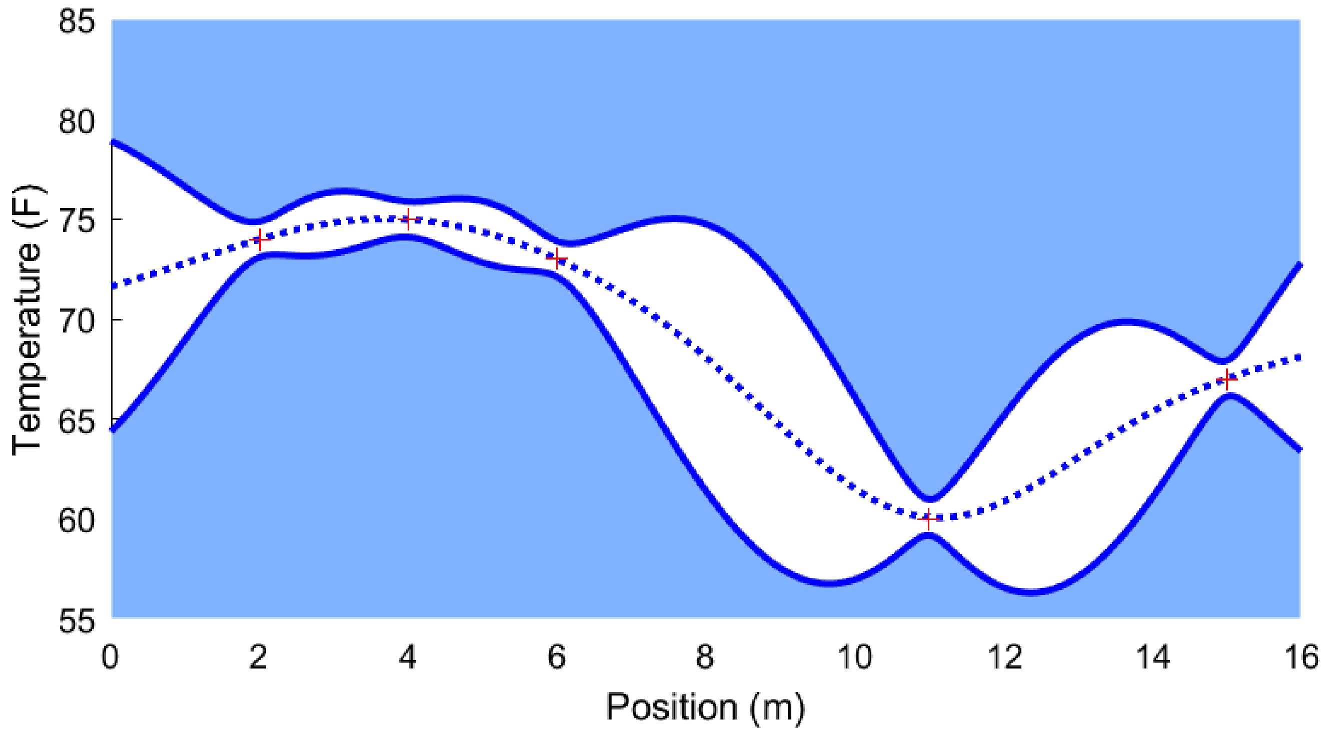
example: ambient temperature field

Gaussian process: $\mathcal{GP}\big(m(x); k(x, x')\big)$

processing information:

local measurements update the field in the surrounding area

Spatially Distributed Phenomena



example: ambient temperature field

Gaussian process: $\mathcal{GP}\big(m(x); k(x, x')\big)$

processing information:

local measurements update the field in the surrounding area

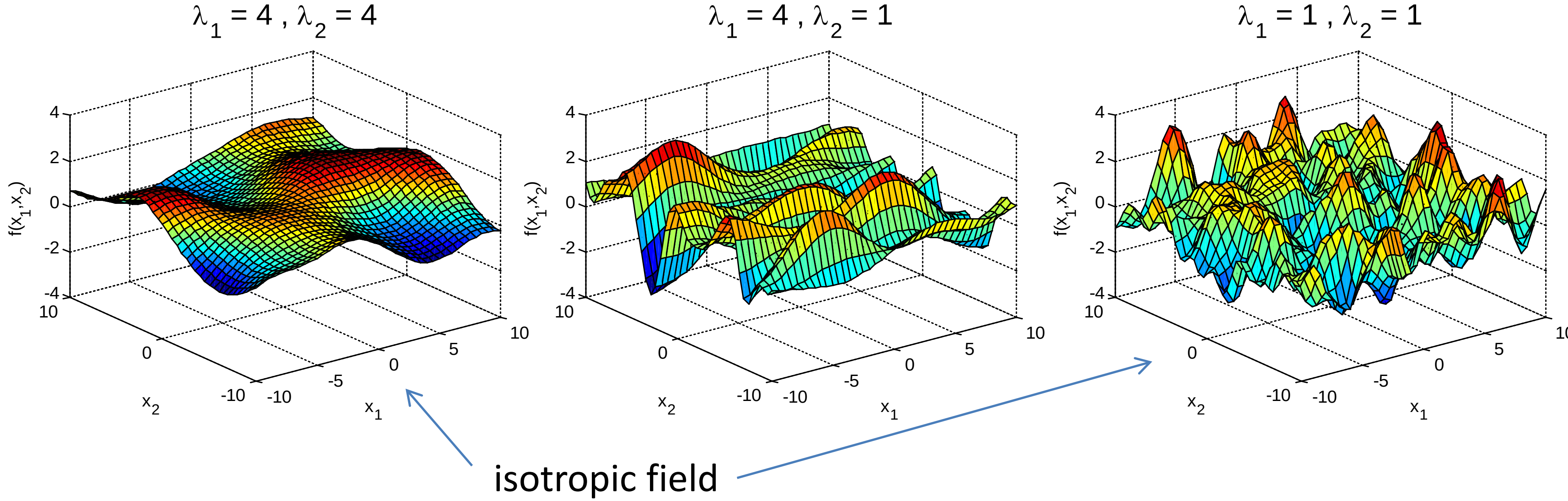
2-D Gaussian Random Field

Physical quantities in 2-d or 3-d: we define a covariance matrix on a grid in higher dimensions.

zero-mean GP with squared exp. cov. funct.

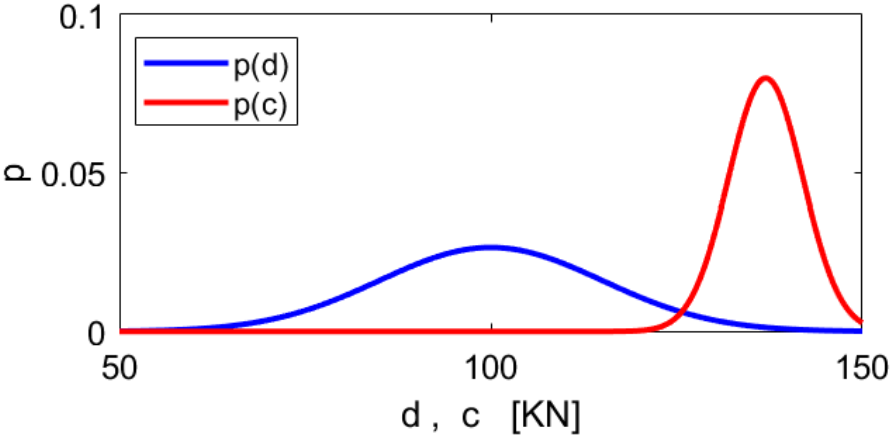
$$f(\mathbf{x}) \sim \mathcal{GP}\left(\mathbf{0}, \sigma_f^2 \exp\left[-\frac{1}{2}\boldsymbol{\Delta}\mathbf{x}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Delta}\mathbf{x}\right]\right)$$

$$\boldsymbol{\Delta}\mathbf{x} = \mathbf{x} - \mathbf{x}' \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$



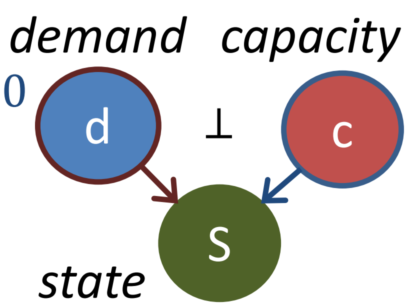
Computation cost grows with number of points, not with number of dimensions.

Gaussian Variables and Decision-Making



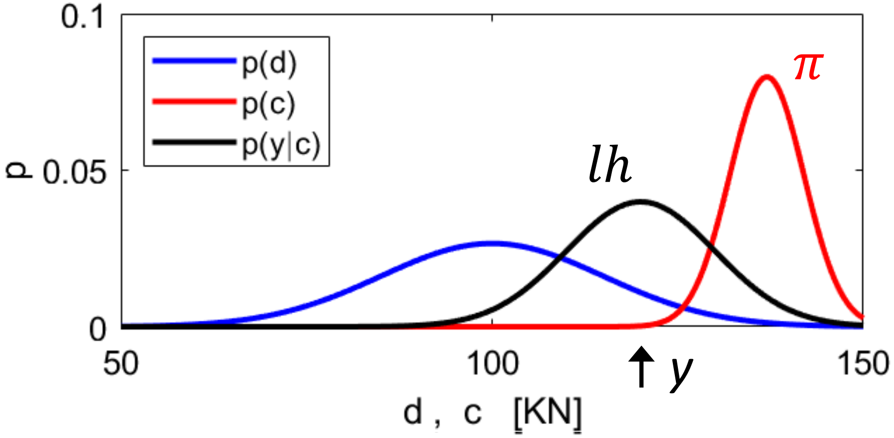
$P_F \quad = 0.96\%$ [Failure iff d > c]

$\mathcal{N}$
$\mu_d = 100$
$\sigma_d = 15$

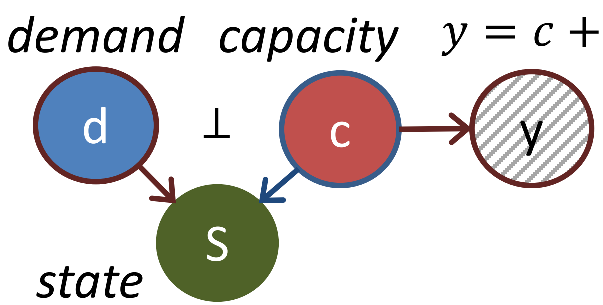


$\mathcal{N}$
$\mu_c = 137$
$\sigma_c = 5$

Gaussian Variables and Decision-Making



$P_F \quad = 0.96\%$ [Failure iff d > c]

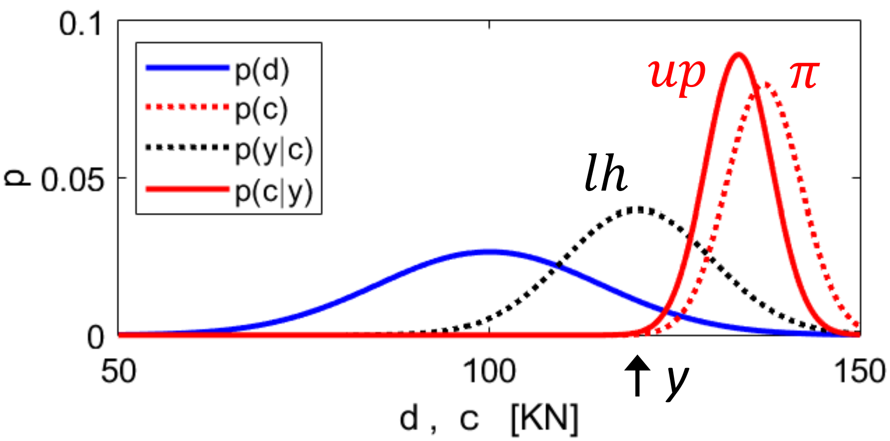


$y = c + \varepsilon$

$\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

$\sigma_\varepsilon = 10$

Gaussian Variables and Decision-Making



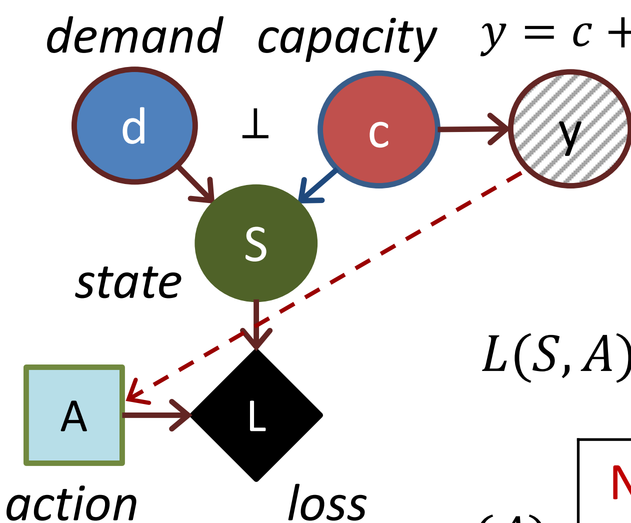
$P_F = 0.96\% \longrightarrow$ **Do Nothing**

$P_{F|y} = 1.55\% \longrightarrow$ **Repair**

Bayes' rule: $up \propto \pi \cdot lh$

optimal expected loss

$$L^* = \min_A L(S,A) = \min\{C_R, P_F C_F\}$$



$y = c + \varepsilon$

$\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$

$\sigma_\varepsilon = 10$

optimal policy

Repair iff $P_F > P^* = \frac{C_R}{C_F} = 1\%$

$L(S,A)$

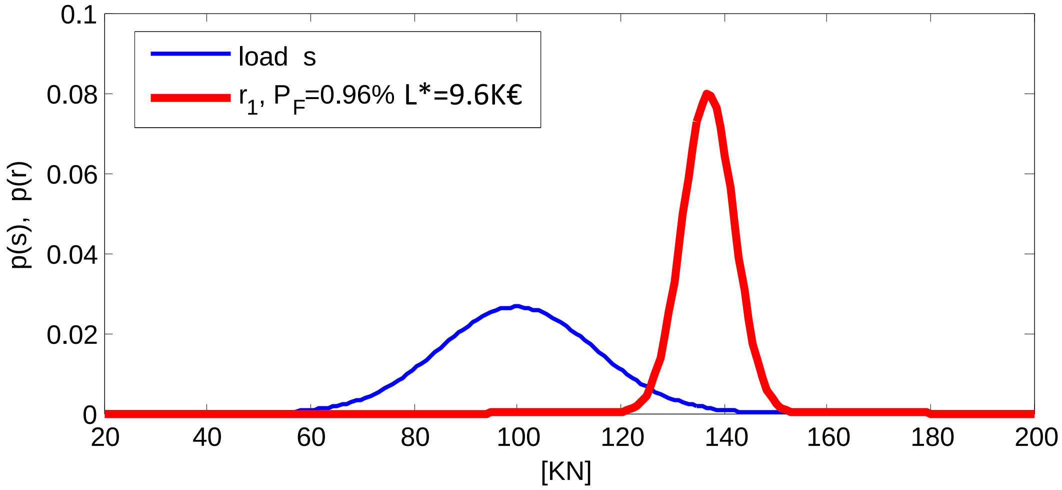
(A) \ (S)	U	F
N	0	C_F
R	C_R	C_R

$C_F = 1\text{M€}$

$C_R = 10\text{K€}$

Value of Information with Gaussian Variables

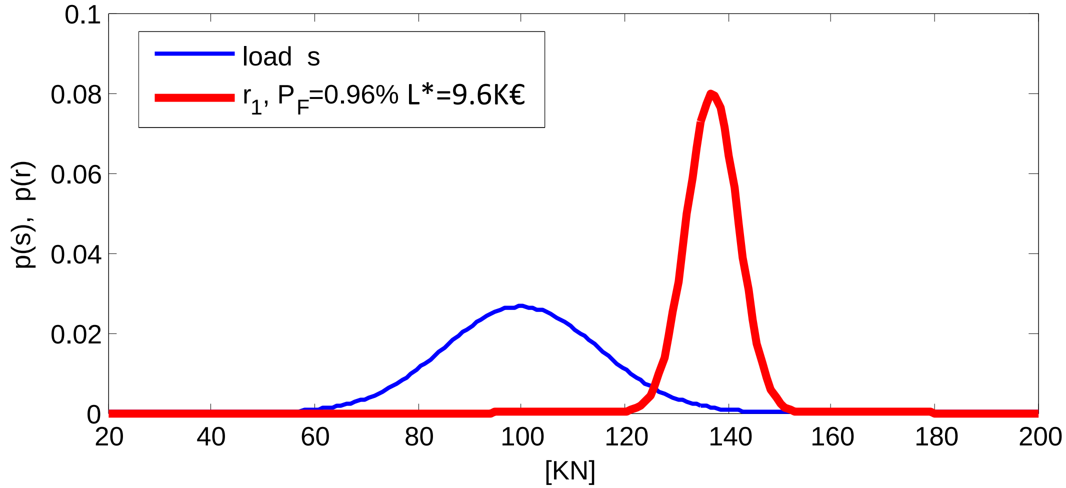
C_F = 1 M€ C_R = 10 K€

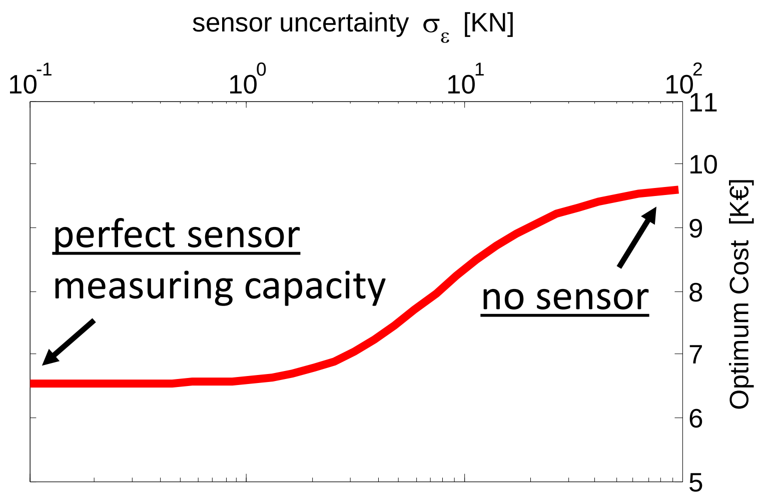


$P_F = 0.96\%$ ➡ A^*: do nothing, L^* = 9.6 K€

Value of Information with Gaussian Variables

C_F = 1 M€ C_R = 10 K€

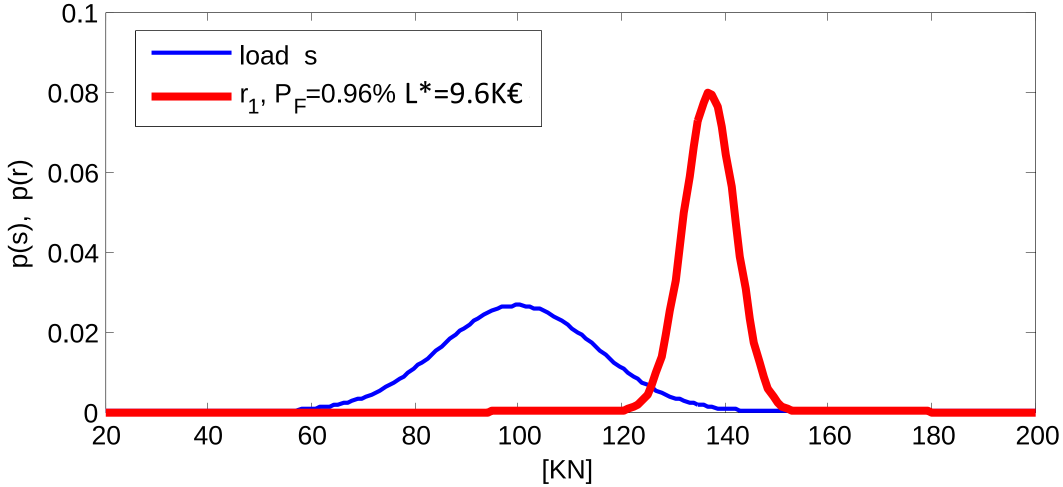




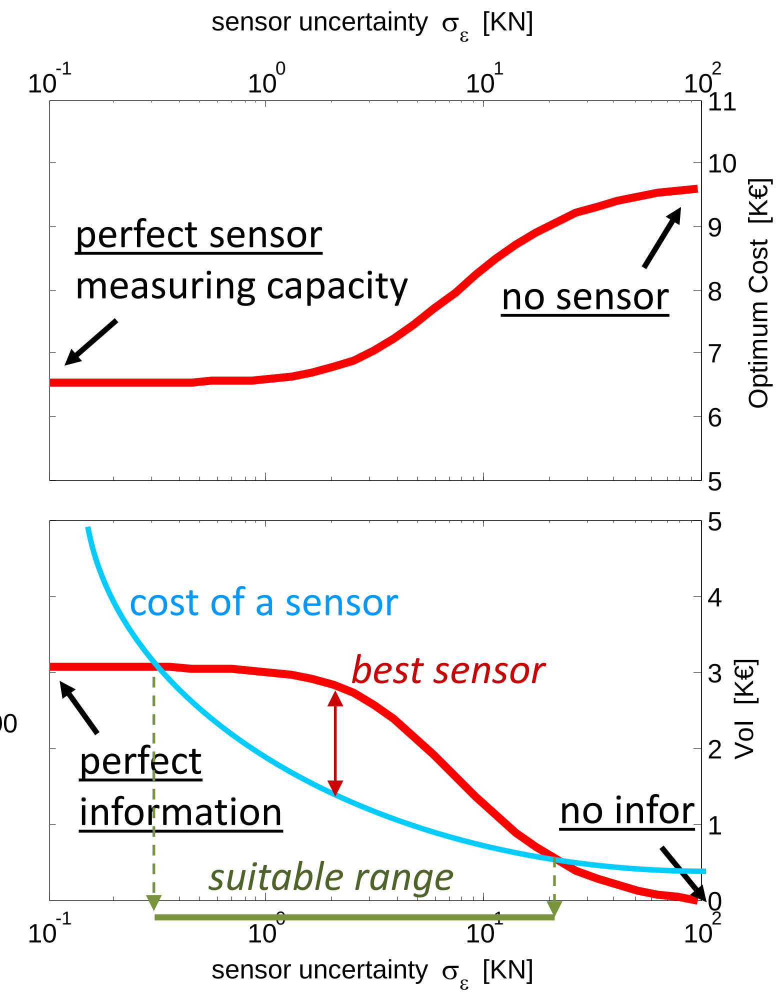
$P_F = 0.96\%$ ➡ A^*: do nothing, L^* = 9.6 K€

Value of Information with Gaussian Variables

C_F = 1 M€ C_R = 10 K€



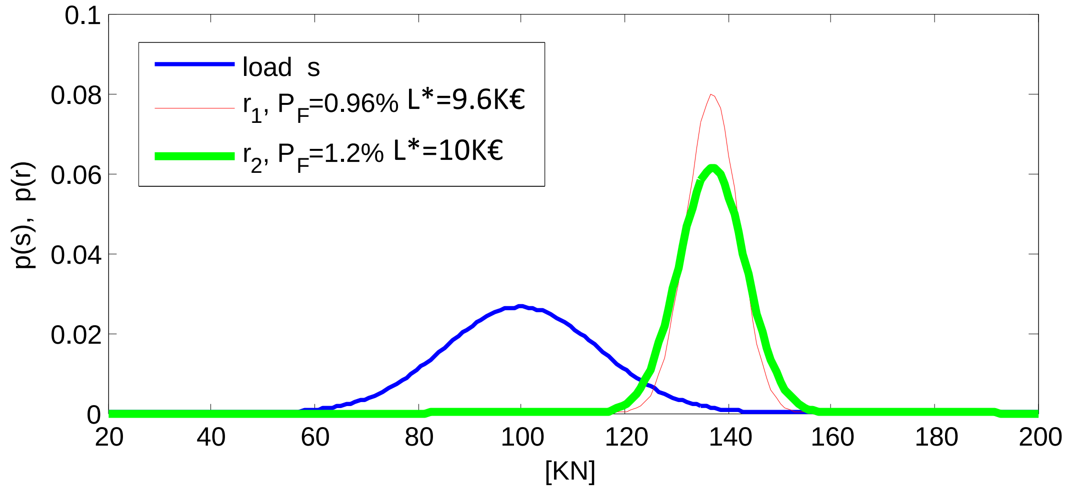
$P_F = 0.96\%$ ➡ A^*: do nothing, L^* = 9.6 K€



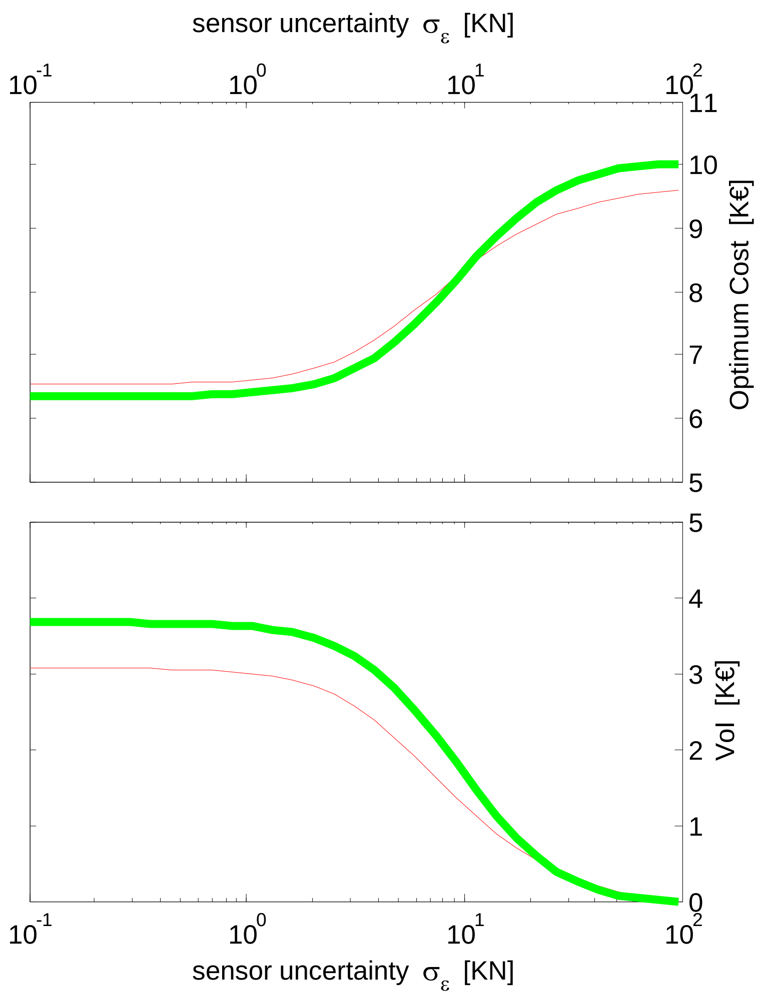
Value of Information with Gaussian Variables

Increasing prior unceraínty

C_F = 1 M€ C_R = 10 K€



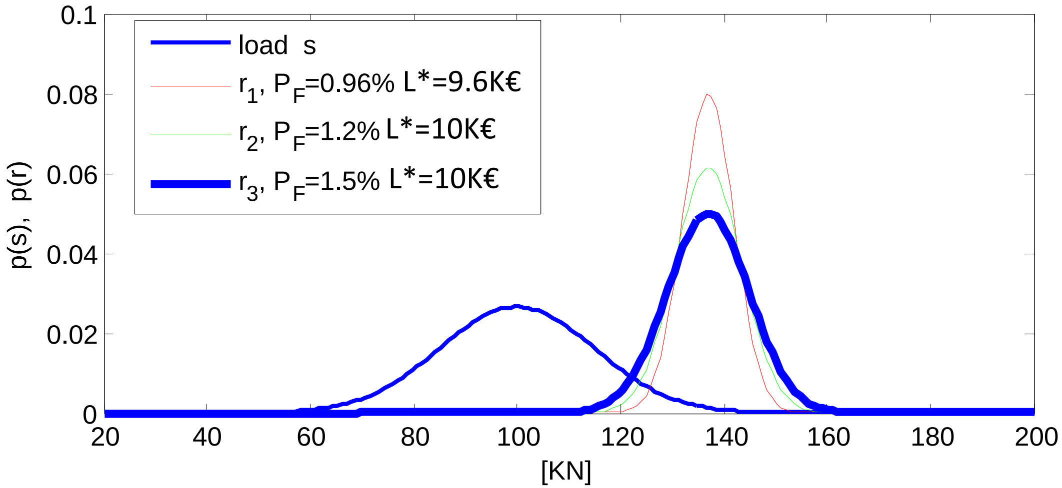
P_F = 1.20% ➡ A^*: repair, L^* = 10 K€

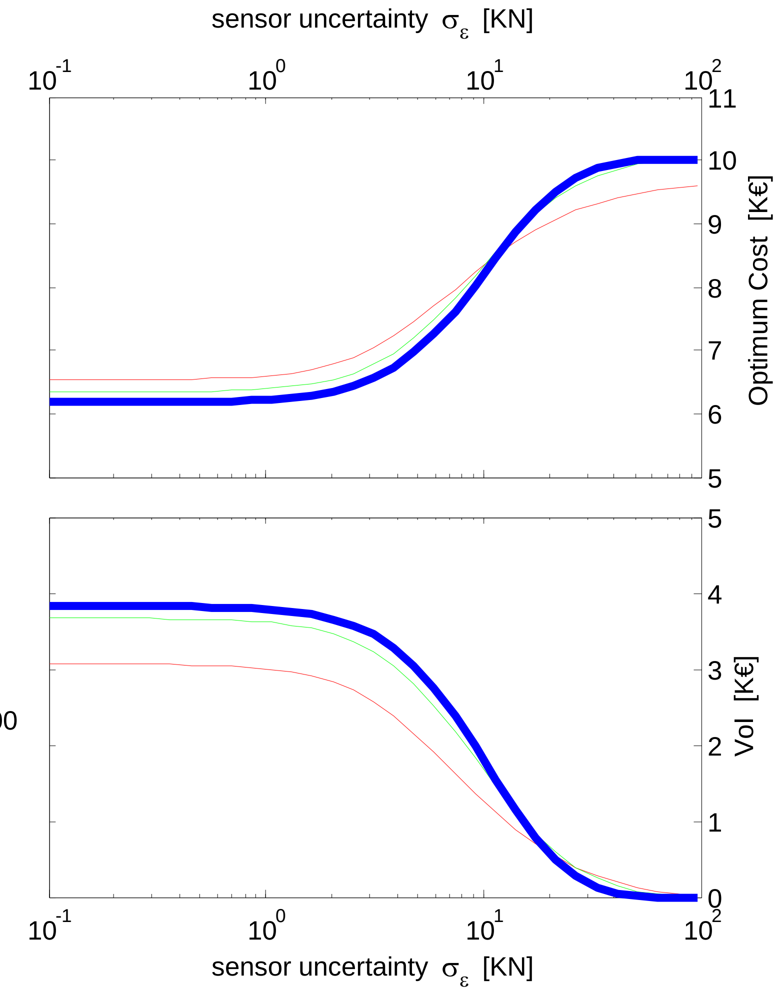


Value of Information with Gaussian Variables

Increasing prior unceraintу

C_F = 1 M€ C_R = 10 K€



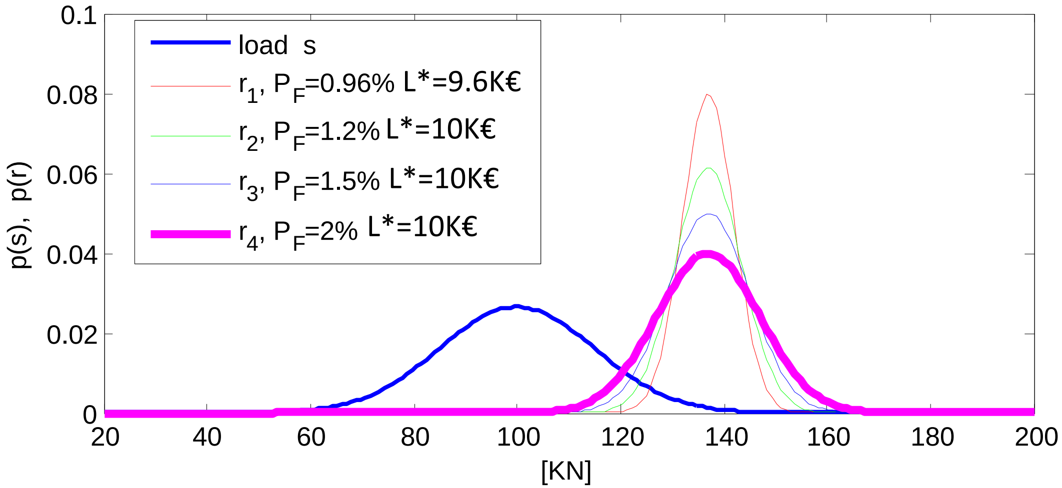


$P_F = 1.50\%$ ➡ *A**: repair, *L** = 10 K€

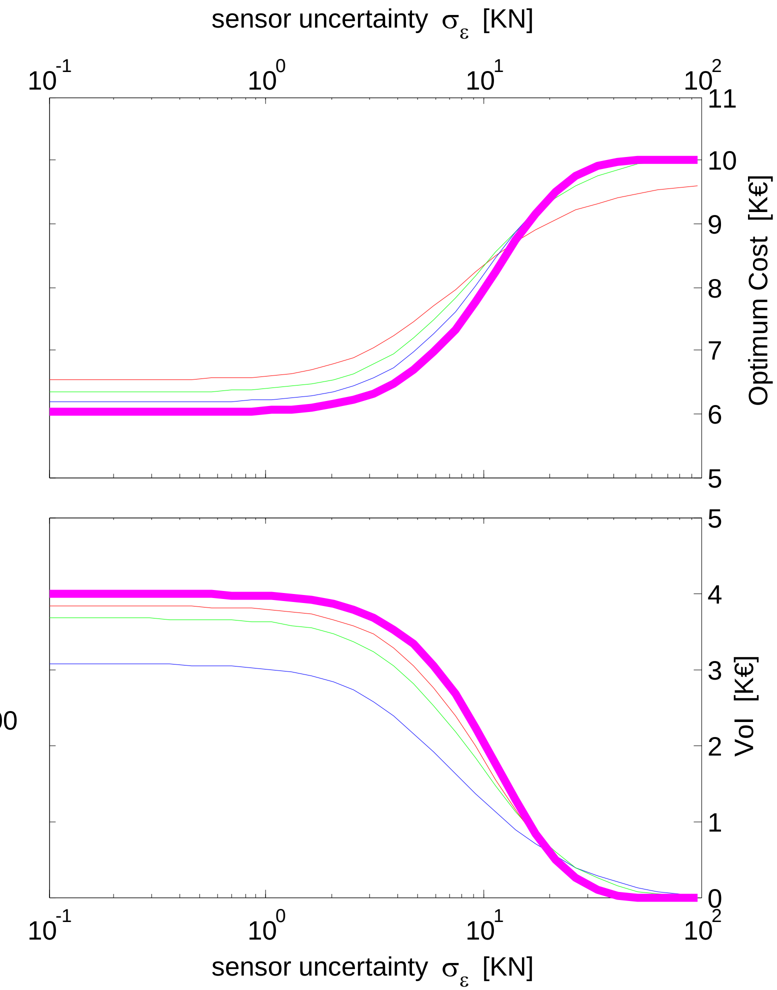
Value of Information with Gaussian Variables

Increasing prior unceraínty

C_F = 1 M€ C_R = 10 K€



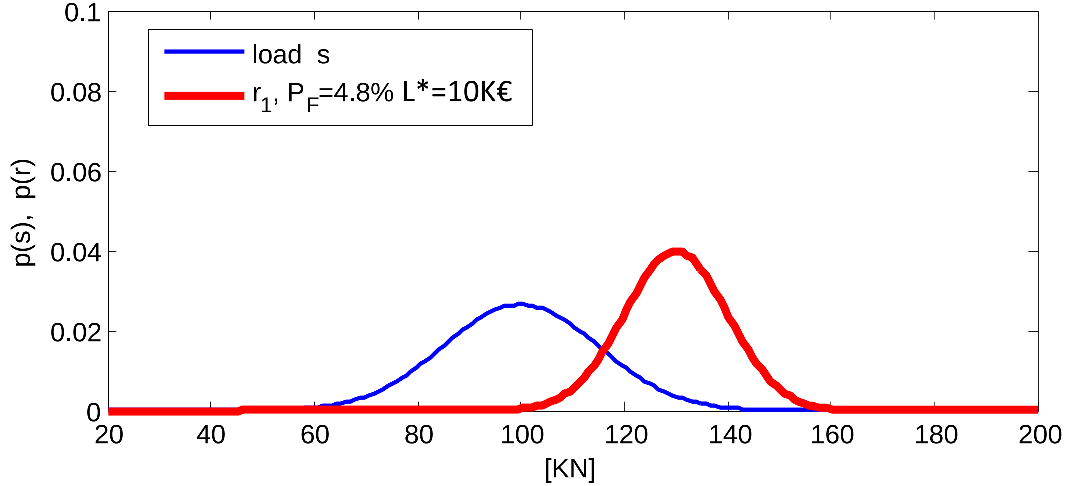
$P_F = 2.05\%$ ➡ A^*: repair, L^* = 10 K€

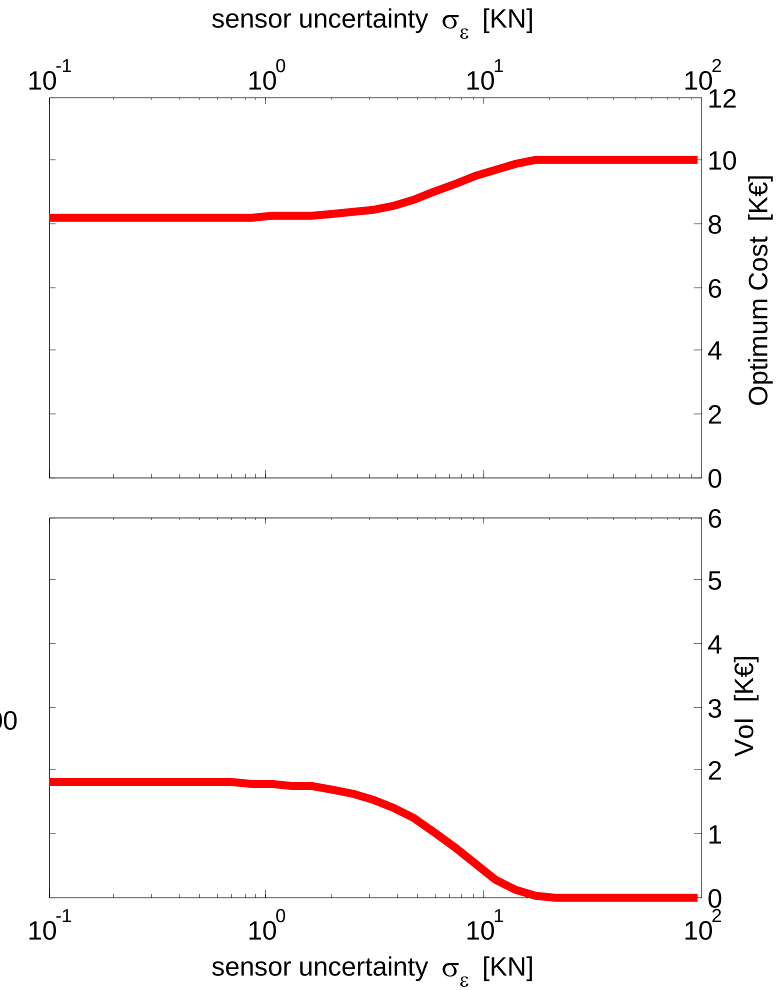


Value of Information with Gaussian Variables

Increasing prior expected capacity

C_F = 1 M€ C_R = 10 K€



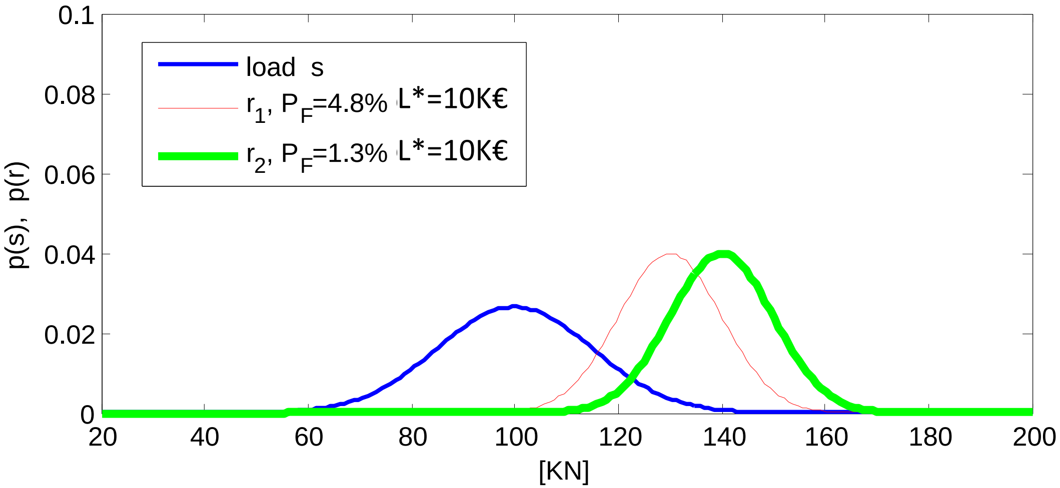


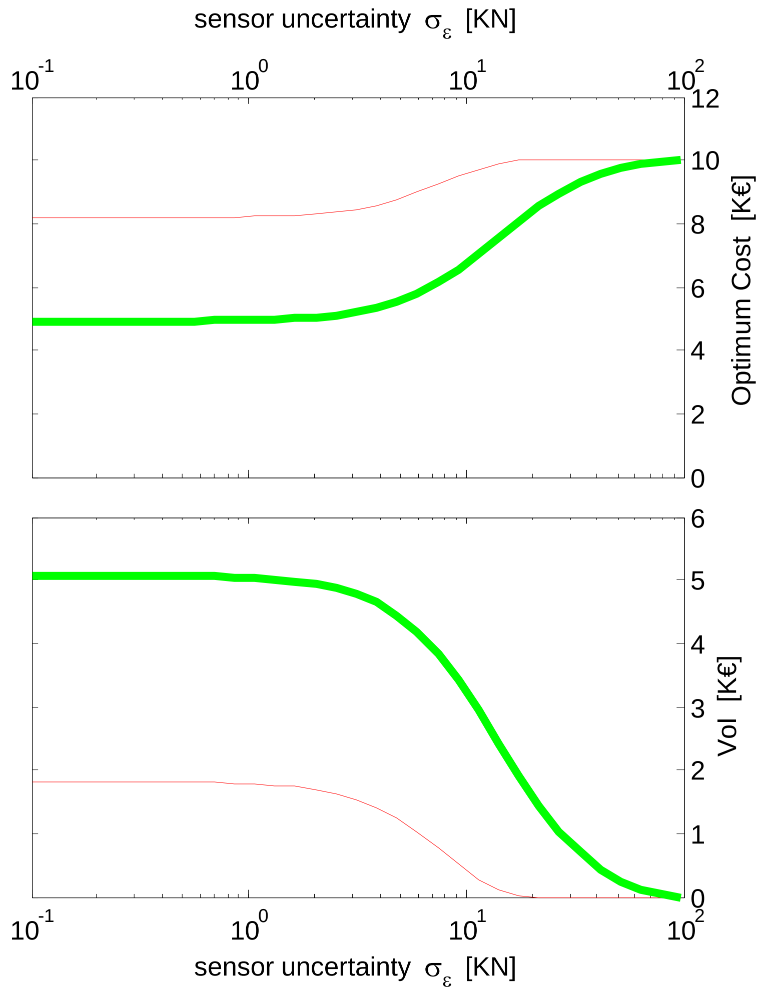
P_F = 4.81% ➡ A^*: repair, L^* = 10 K€

Value of Information with Gaussian Variables

Increasing prior expected capacity

C_F = 1 M€ $\qquad$ C_R = 10 K€



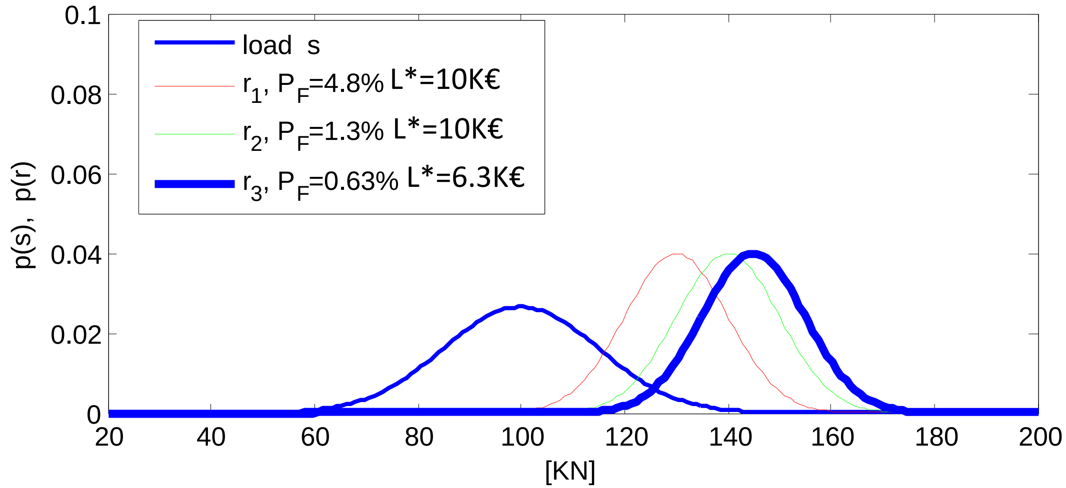


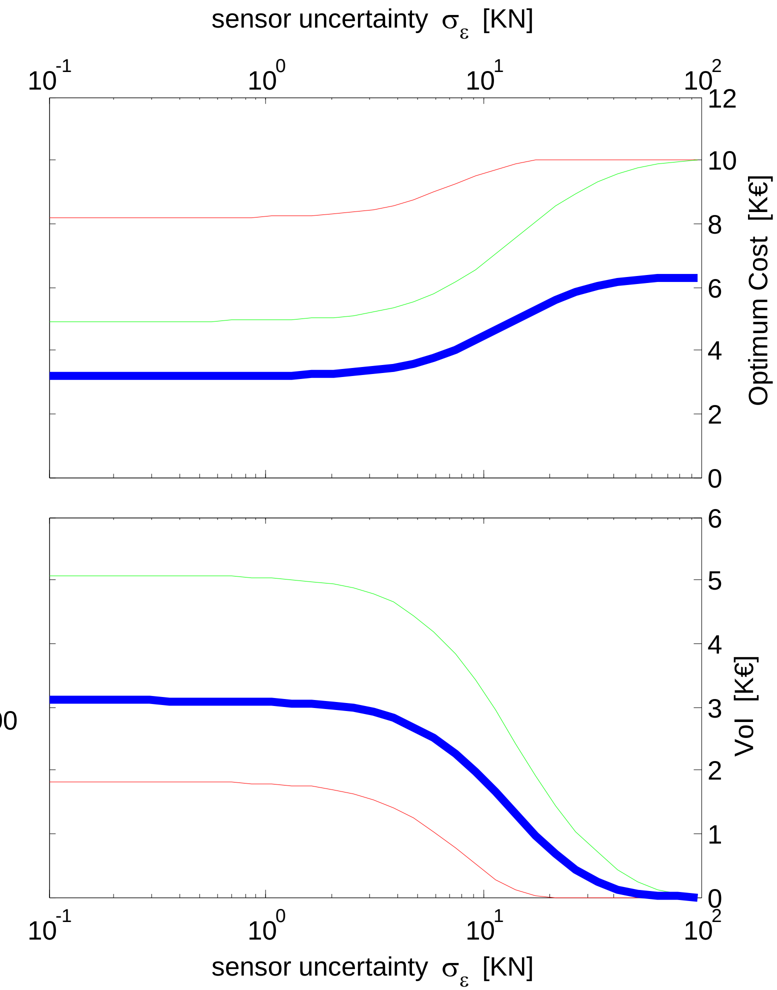
$P_F = 1.28\%$ ➡ A^*: repair, L^* = 10 K€

Value of Information with Gaussian Variables

Increasing prior expected capacity

C_F = 1 M€ C_R = 10 K€



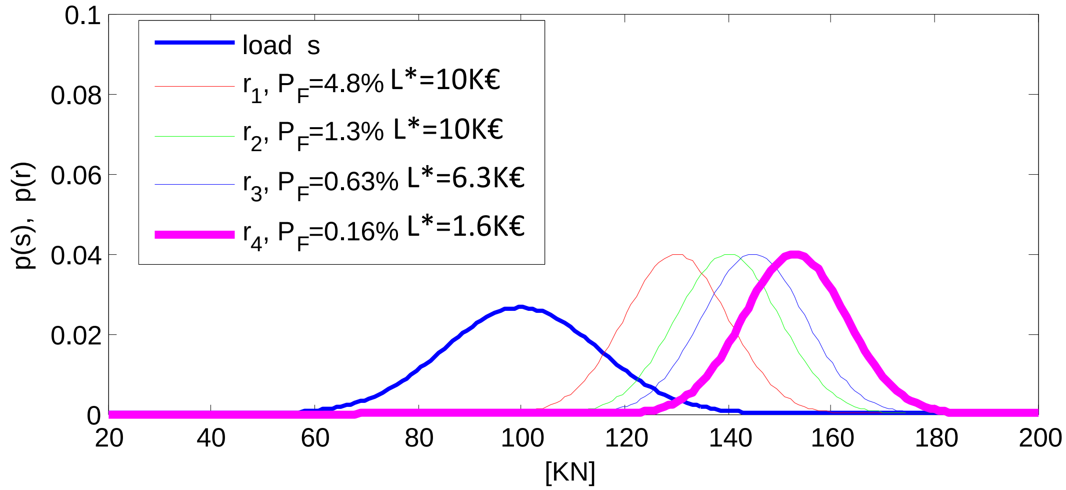


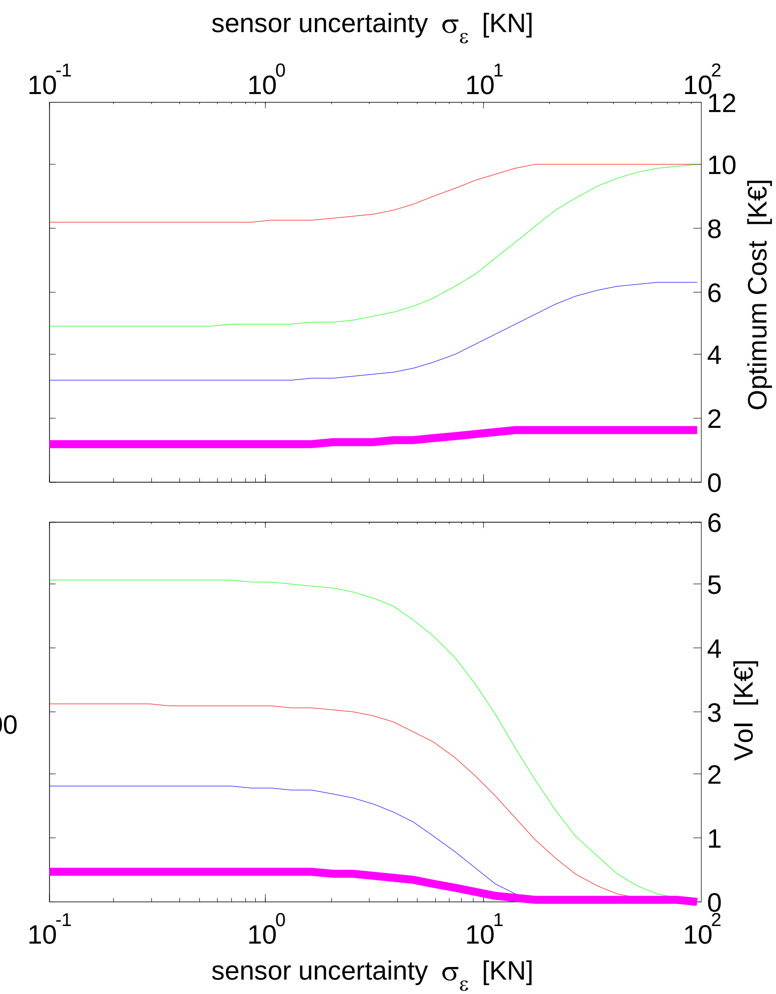
$P_F = 0.63\%$ ➡ A^*: do nothing, L^* = 6.3 K€

Value of Information with Gaussian Variables

Increasing prior expected capacity

C_F = 1 M€ C_R = 10 K€

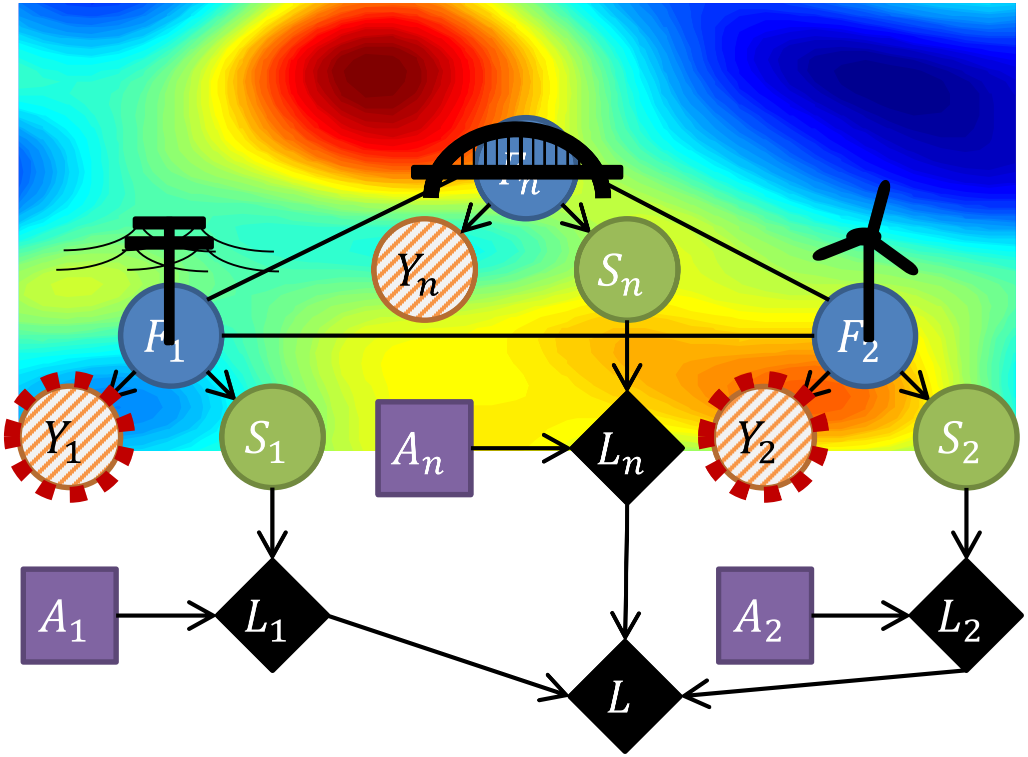




$P_F = 0.16\%$ ➡ A^*: do nothing, L^* = 1.6 K€

Value of Information in Gaussian Random Fields

A model for system performance and decision-making combines the probabilistic random field model for the spatially distributed system (F) with models for observations (Y), components states (S), managing actions (A), and losses (L).



Random Field Models
(e.g. event, load, capacity)

Measurement Models
(e.g. strain gauge)

Structural & Limit State Models
(e.g. FEM analysis)

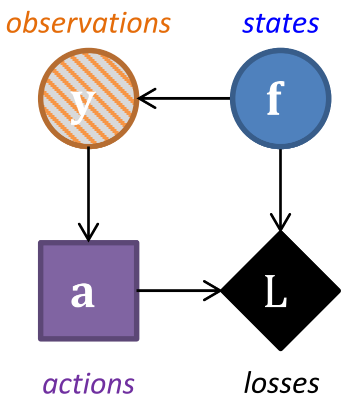
Decision Models
(e.g. repair, replace)

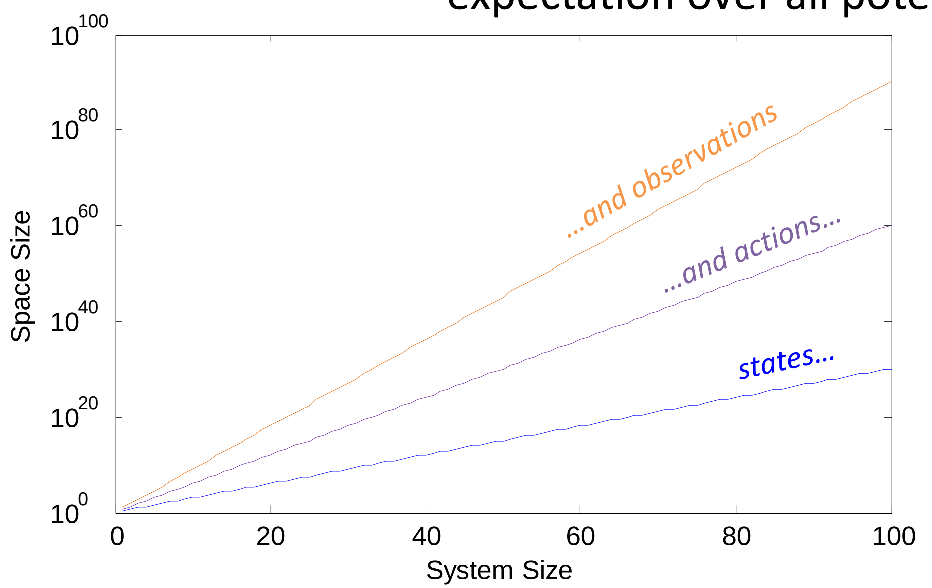
Consequence Models
(e.g. failure, loss of function)

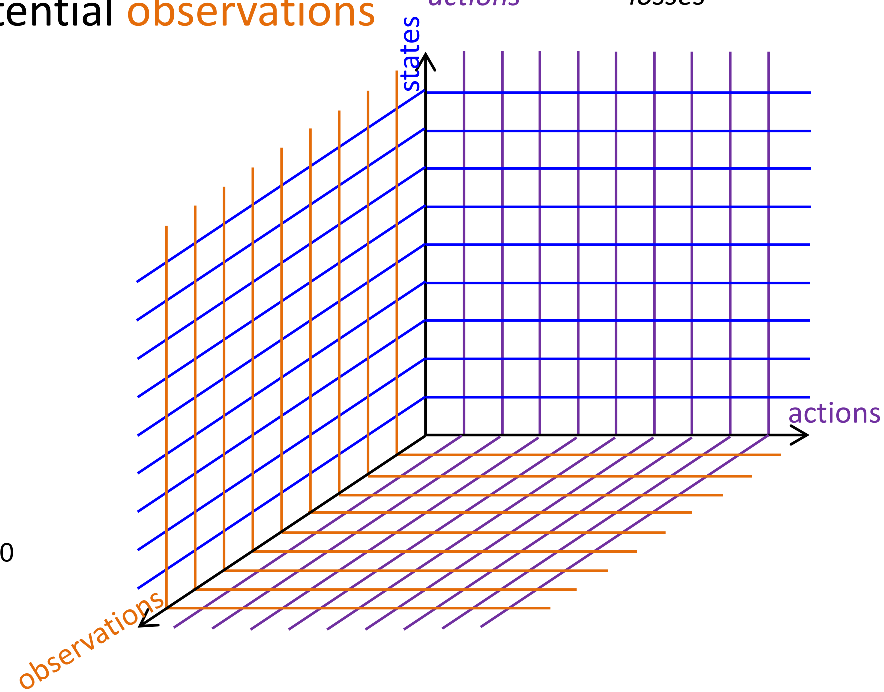
Measurement Optimization
(e.g. VoI Analysis)

Computational Complexity in Large Systems

$$\mathbb{E}L(Y) = \underbrace{\mathbb{E}_Y \underbrace{\min_A \underbrace{\mathbb{E}_{F|\mathbf{y}} L(\mathbf{f},\mathbf{a})}_{\text{expectation over all joint states}}}_{\text{minimization over all joint actions}}}_{\text{expectation over all potential observations}}$$

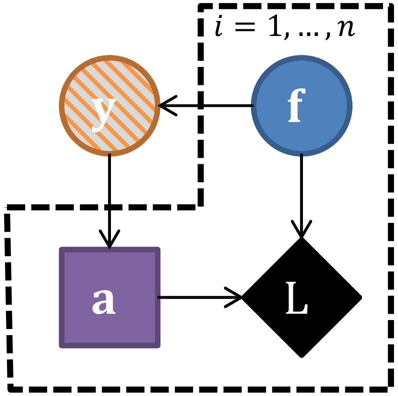






Cumulative System Assumption

Local loss is a function of local actions and local states



Malings, C., and Pozzi, M. , 2016. "Conditional Entropy and Value of Information Metrics for Optimal Sensing in Infrastructure Systems", *Structural Safety*, 60: 77-90.

$$\mathrm{L}(\mathbf{f},\mathbf{a}) = \sum_{i=1}^{n} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \min_A \mathbb{E}_{F|\mathbf{y}} \sum_{i=1}^{n} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \min_A \sum_{i=1}^{n} \mathbb{E}_{F_i|\mathbf{y}} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \sum_{i=1}^{n} \min_{A_i} \mathbb{E}_{F_i|\mathbf{y}} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(Y) = \sum_{i=1}^{n} \mathbb{E}_Y \min_{A_i} \mathbb{E}_{F_i|\mathbf{y}} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(\emptyset) = \sum_{i=1}^{n} \mathbb{E}_\emptyset \min_{A_i} \mathbb{E}_{F_i|\emptyset} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(\emptyset) = \sum_{i=1}^{n} \min_{A_i} \mathbb{E}_{F_i} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

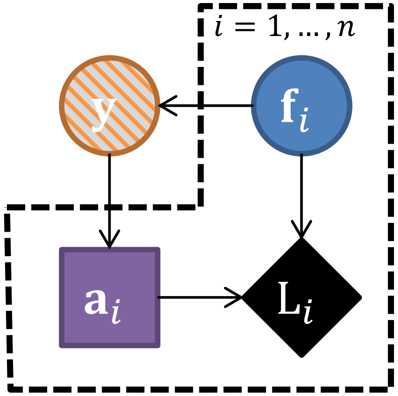
$$\mathrm{VoI}_i(Y) = \min_{A_i} \mathbb{E}_{F_i} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i) - \mathbb{E}_Y \min_{A_i} \mathbb{E}_{F_i|\mathbf{y}} \mathrm{L}_i(\mathbf{f}_i,\mathbf{a}_i)$$

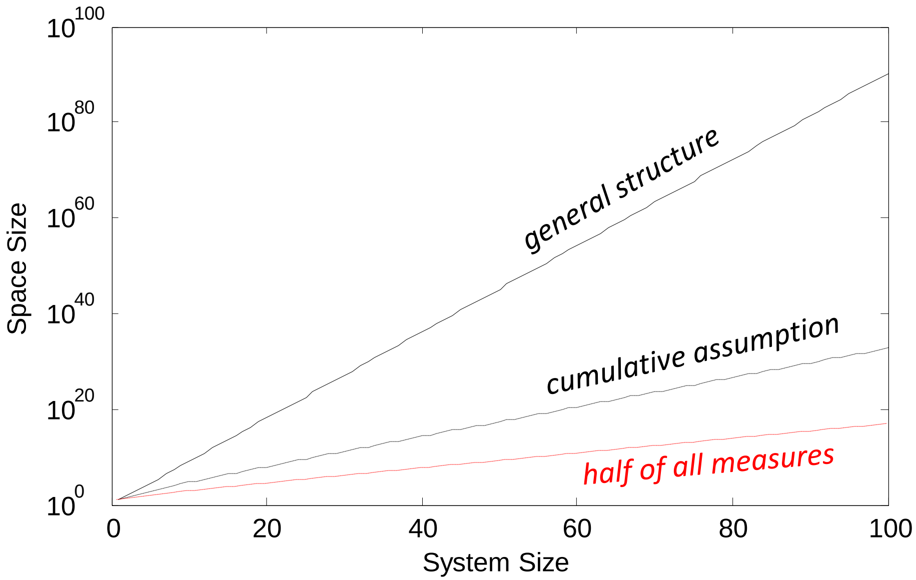
$$\mathrm{VoI}(Y) = \sum_{i=1}^{n} \mathrm{VoI}_i(Y)$$

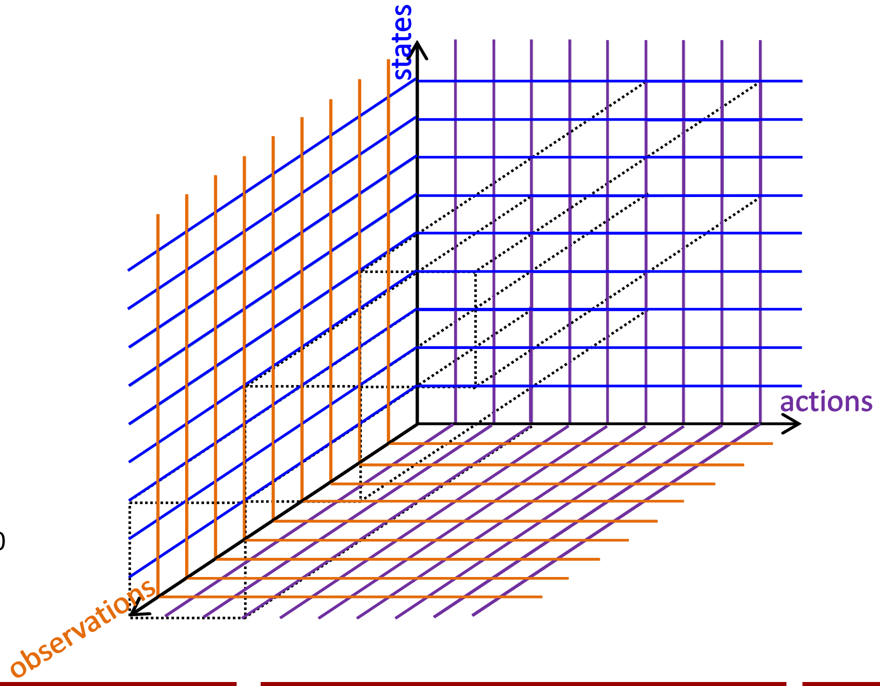
Cumulative System Assumption

$$\mathrm{L}(\mathbf{f}, \mathbf{a}) = \sum_{i=1}^{n} \mathrm{L}_i(\mathbf{f}_i, \mathbf{a}_i)$$

$$\mathbb{E}\mathrm{L}(Y) = \sum_{i=1}^{n} \mathbb{E}_Y \min_{A_i} \mathbb{E}_{F_i|\mathbf{y}} \mathrm{L}_i(\mathbf{f}_i, \mathbf{a}_i)$$







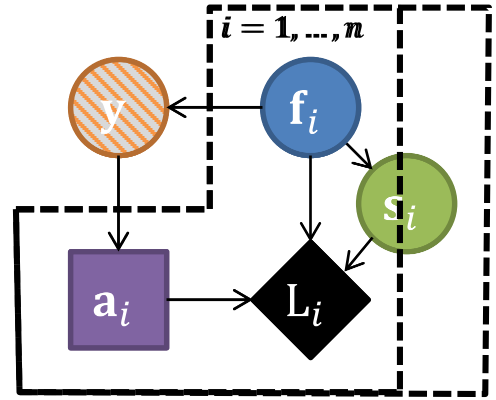
Value of Information with Gaussian Random Variables

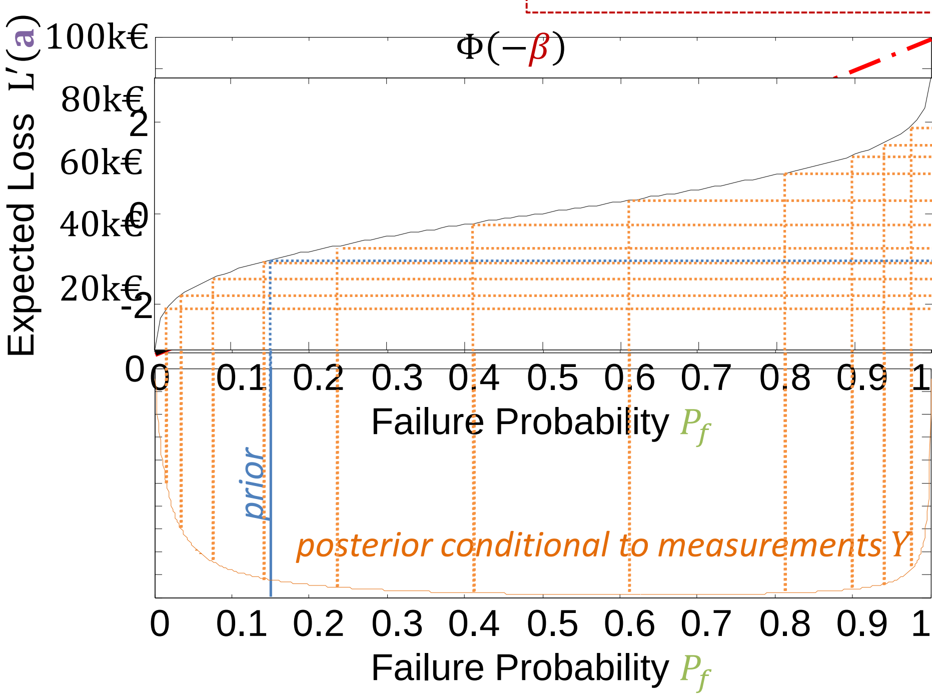
$$\mathbf{f} \sim \mathcal{N}(\boldsymbol{\mu}_F, \boldsymbol{\Sigma}_F)$$

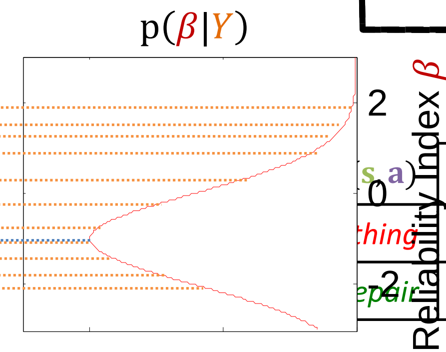
$$\mathbf{y} = \mathbf{R}_Y \mathbf{f} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\mu}_\epsilon, \boldsymbol{\Sigma}_\epsilon) \qquad \mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y)$$

$$\mathbf{s} = \mathbb{I}(\boldsymbol{\Xi}\mathbf{f} + \xi \geq 0)$$

Malings, C., and Pozzi, M. , 2016. "Conditional Entropy and Value of Information Metrics for Optimal Sensing in Infrastructure Systems", *Structural Safety*, 60: 77-90.







(s, a)	*operational*	*failed*
thing	0	100k€
epair	25k€	25k€

State ($\mathbf{s}$)

$$\mathbb{E}\mathrm{L}(\emptyset) = \mathrm{L}^*\left(P_{f,\mathrm{prior}}\right)$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \mathrm{L}^*\left(P_f|\mathbf{y}\right)$$

$$\beta = -\Phi^{-1}\left(P_f\right)$$

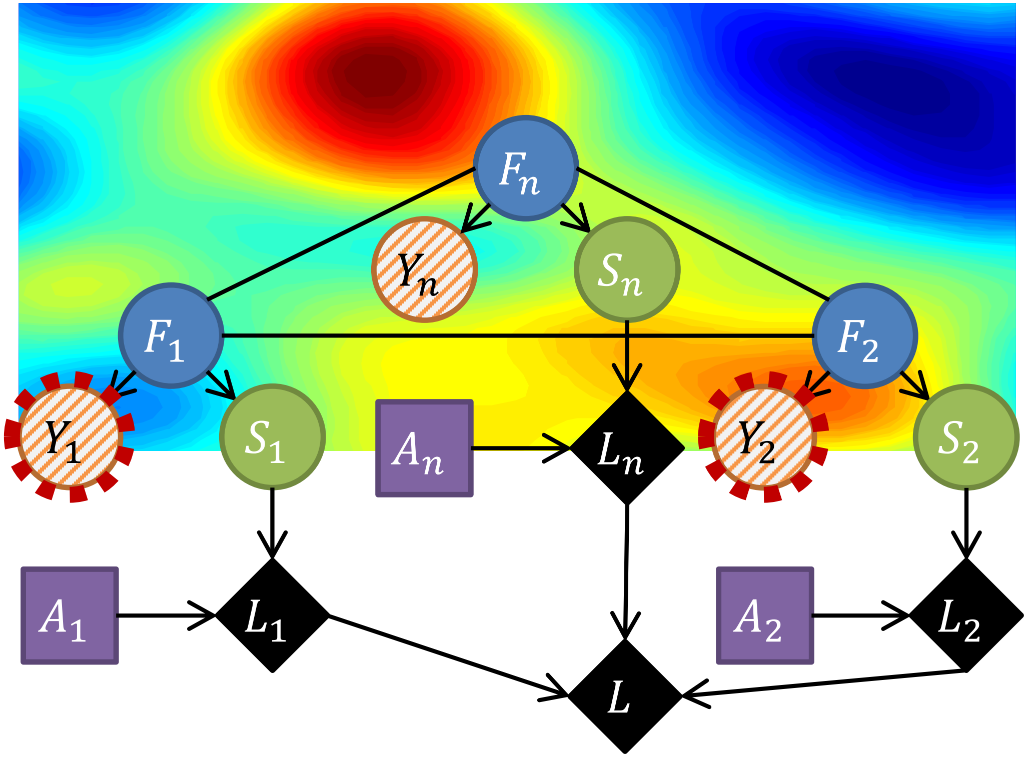
$$\beta|Y \sim \mathcal{N}(\mu_{\beta|Y}, \sigma_{\beta|Y}) \text{ if } \mathbf{f}, \mathbf{y} \sim \mathcal{N}$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_{\beta|Y} \mathrm{L}^{**}(\beta)$$

Value of Information in Gaussian Random Fields

A model for system performance and decision-making combines the probabilistic random field model for the spatially distributed system (F) with models for observations (Y), components states (S), managing actions (A), and losses (L).

Under the cumulative system assumption and in Gaussian random fields, value of information can be efficiently evaluated to compare potential sensing schemes.



Observation Model:

$$\mathbf{y} = \mathbf{R}_Y \mathbf{f} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$$

State Model:

$$\mathbf{s} = \mathbb{I}\left(\boldsymbol{\Xi}^{\mathrm{T}} \mathbf{f} \geq \mathbf{0}\right)$$

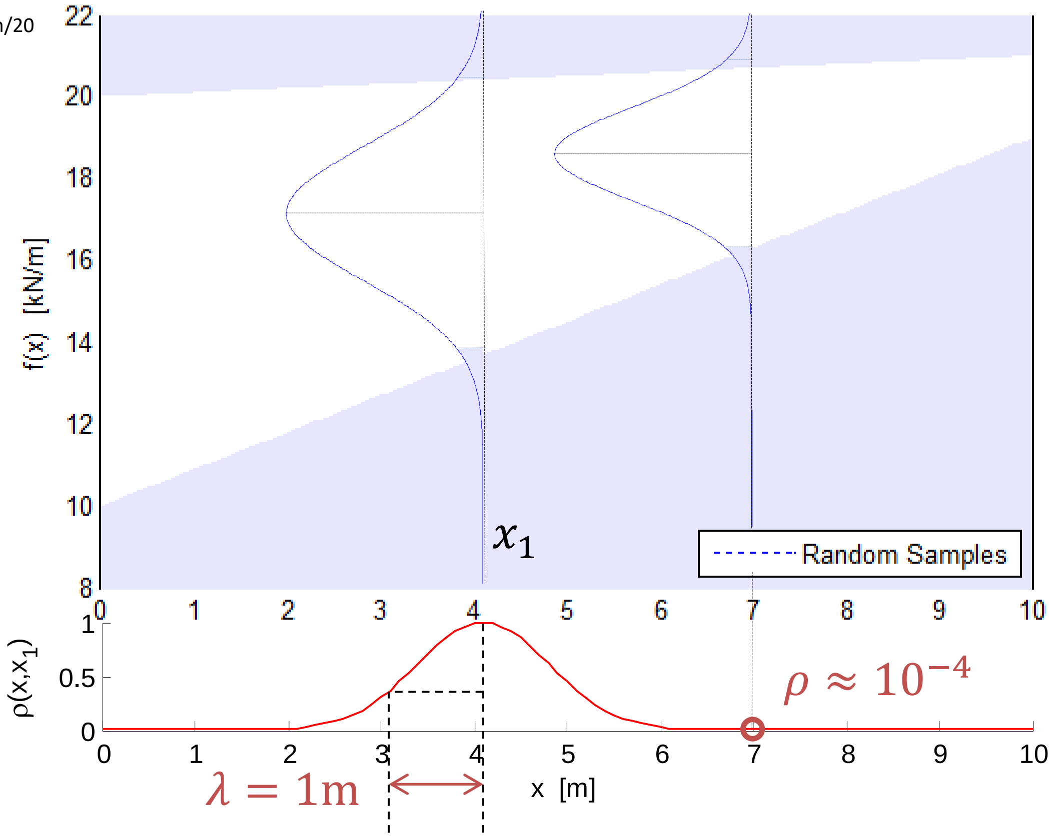
Loss Model:

$$L(\mathbf{s}, \mathbf{a}) = \sum_{i=1}^{n} L_i(s_i, a_i)$$

Gaussian Loading Model

http://threedayslong.blogspot.com/2011/02/one-week-ago-pic.html

http://murmurcreekobservatory.blogspot.com/2014/01/snow-cold-repeat-and-our-propane-is-12.html



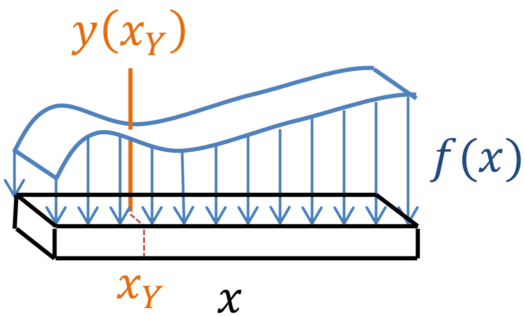
Gaussian Random Field

- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$

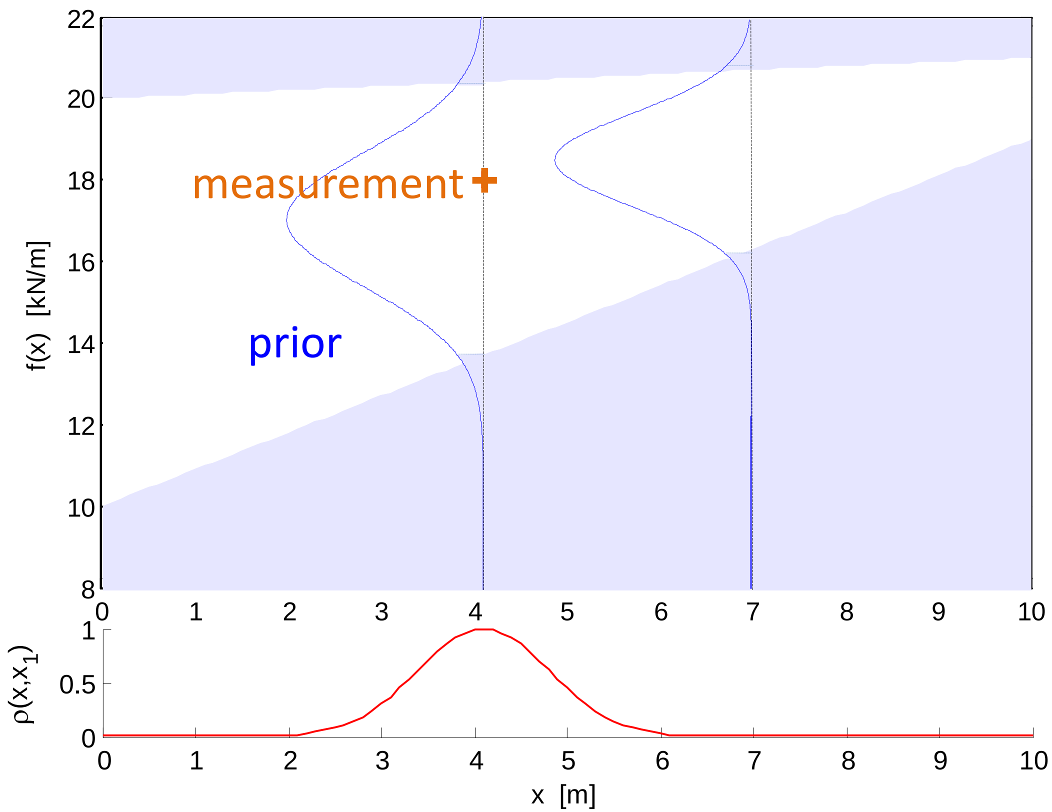
Probabilistic Inference

$$f(x) \sim \mathcal{GP}\left(\mu(x), K(x, x')\right)$$

$$y(x_Y) \sim \mathcal{N}\left(\mu(x_Y), \sigma^2(x_Y) + \epsilon_Y{}^2\right)$$



$$\boldsymbol{f} \sim \mathcal{N}\left(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f\right)$$

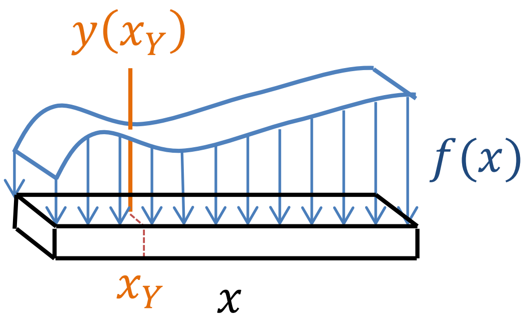


Gaussian Random Field

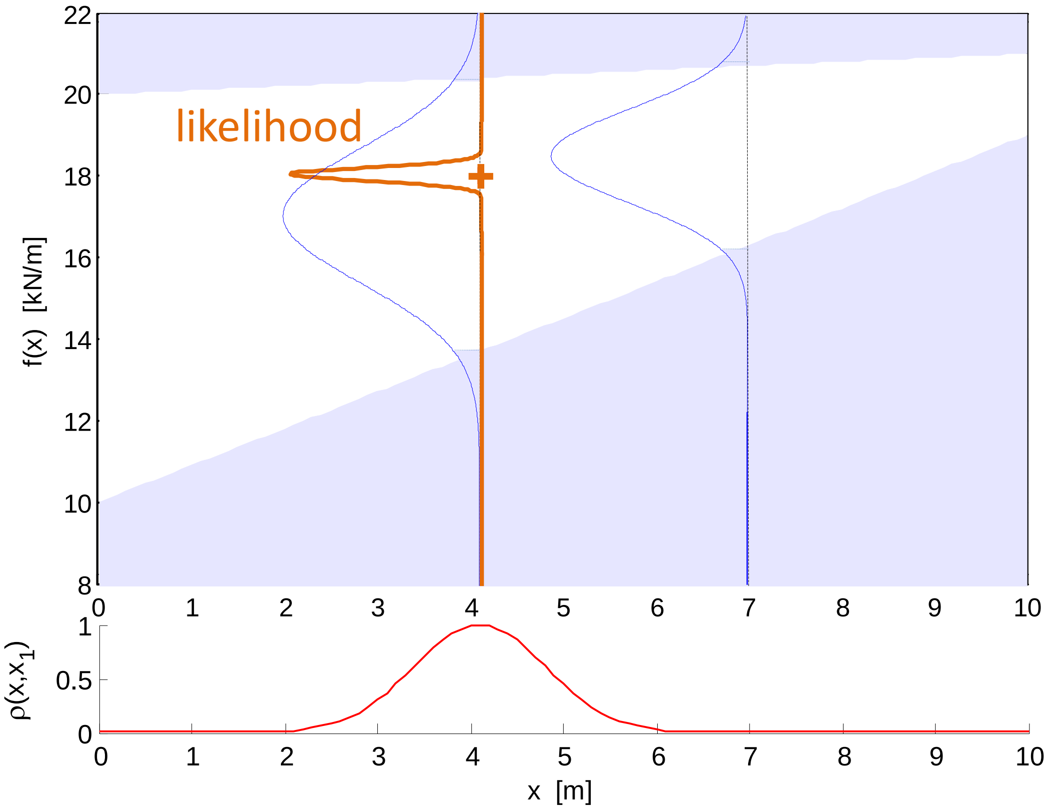
- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$
- Updating – single snow depth measure

$$f(x) \sim \mathcal{GP}\big(\mu(x), K(x,x')\big)$$

$$y(x_Y) \sim \mathcal{N}\big(\mu(x_Y), \sigma^2(x_Y) + \epsilon_Y{}^2\big)$$



$$\boldsymbol{f} \sim \mathcal{N}\big(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f\big)$$



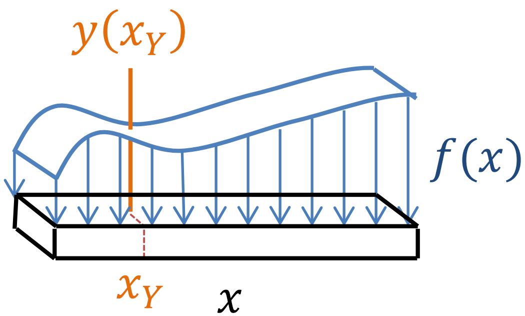
Gaussian Random Field

- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$
- Updating – single snow depth measure

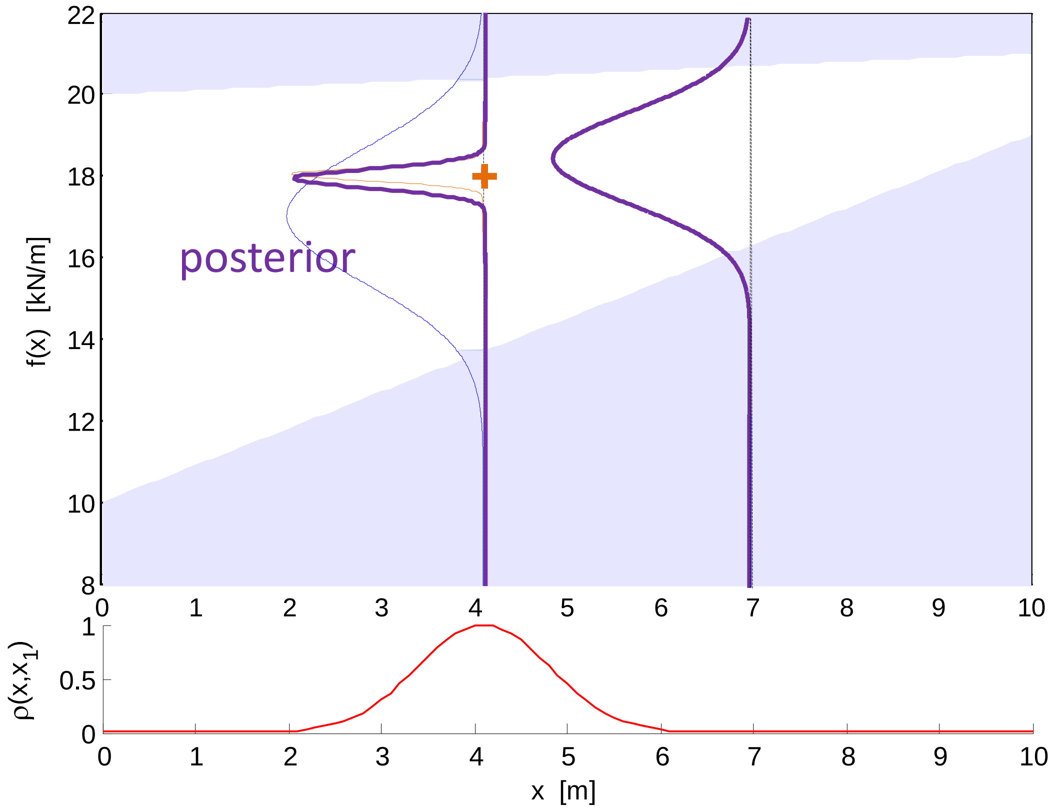
Probabilistic Inference

$$f(x) \sim \mathcal{GP}\big(\mu(x), K(x,x')\big)$$

$$y(x_Y) \sim \mathcal{N}\big(\mu(x_Y), \sigma^2(x_Y) + {\epsilon_Y}^2\big)$$



$$\boldsymbol{f}|y \sim \mathcal{N}\big(\boldsymbol{\mu}_{f|y}, \boldsymbol{\Sigma}_{f|Y}\big)$$



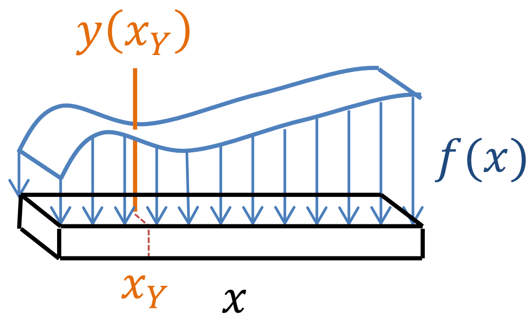
Gaussian Random Field

- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$
- Updating – single snow depth measure

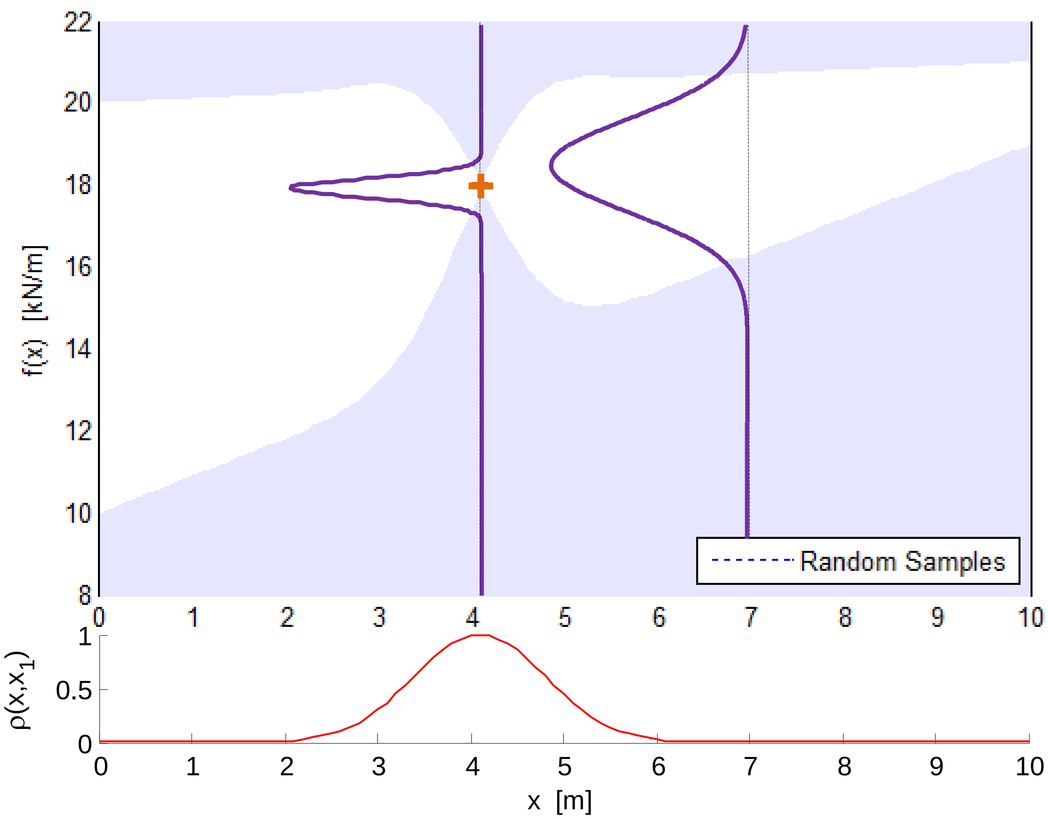
Probabilistic Inference

$$f(x) \sim \mathcal{GP}\big(\mu(x), K(x, x')\big)$$

$$y(x_Y) \sim \mathcal{N}\big(\mu(x_Y), \sigma^2(x_Y) + \epsilon_Y{}^2\big)$$



$$\boldsymbol{f}|y \sim \mathcal{N}\big(\boldsymbol{\mu}_{f|y}, \boldsymbol{\Sigma}_{f|Y}\big)$$



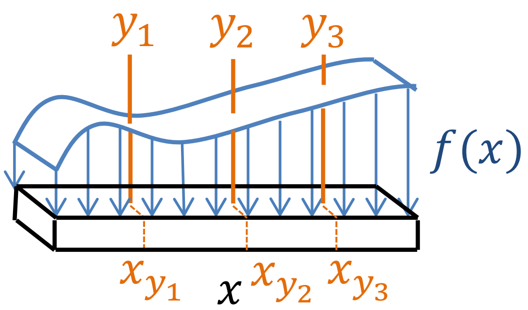
Gaussian Random Field

- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$
- Updating – single snow depth measure

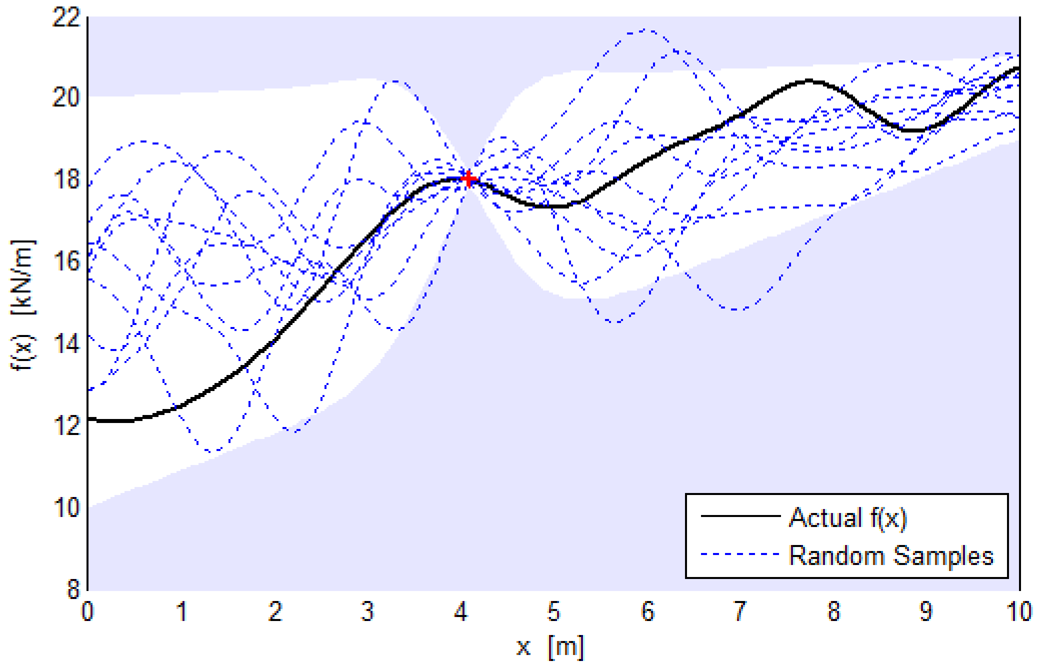
Probabilistic Inference

$$f(x) \sim \mathcal{GP}\big(\mu(x), K(x, x')\big)$$

$$\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{\mu}(X_Y), \boldsymbol{\Sigma}_{f(X_Y)} + \boldsymbol{\Sigma}_{\epsilon}\big)$$



$$\boldsymbol{f}|\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{\mu}_{f|y}, \boldsymbol{\Sigma}_{f|Y}\big)$$



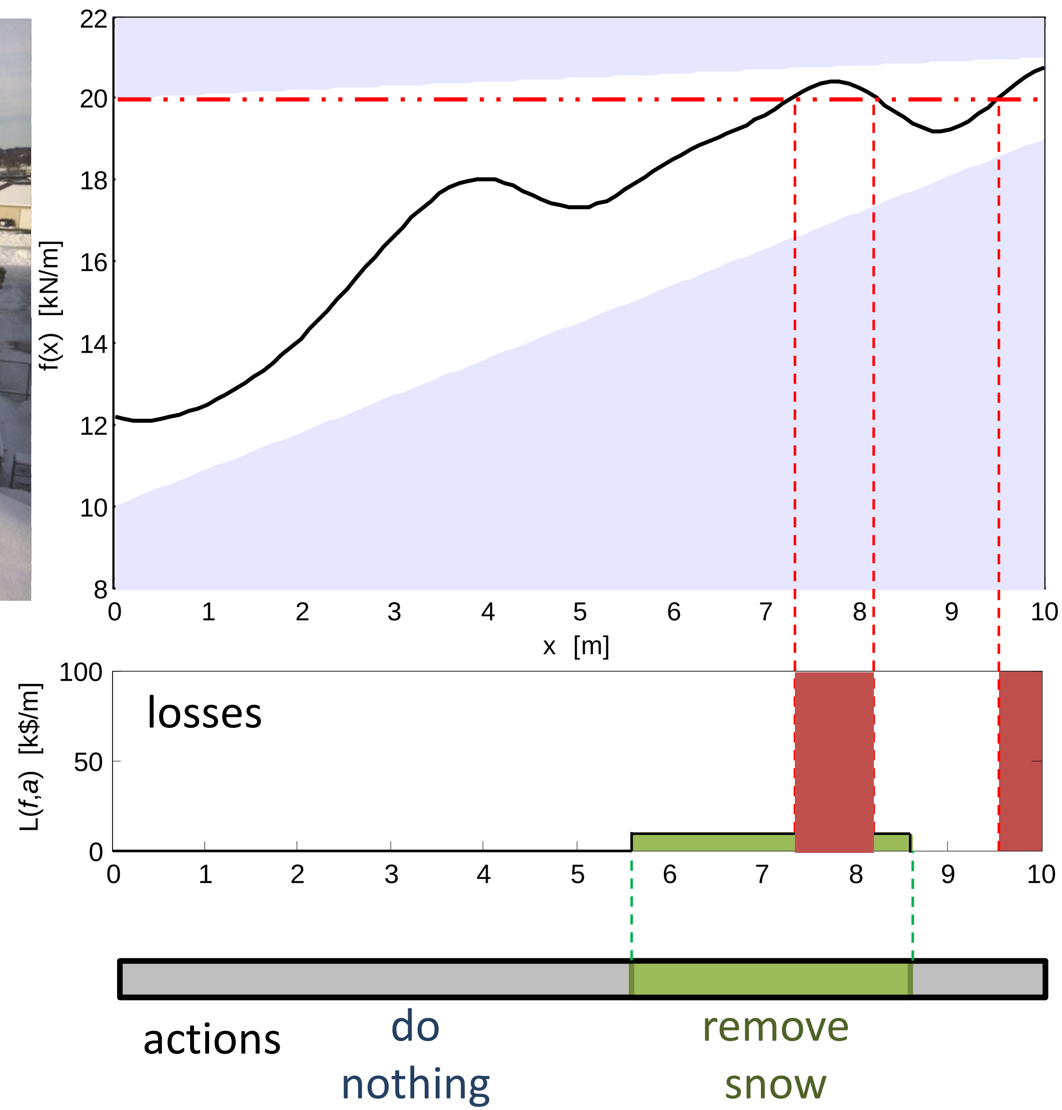
Gaussian Random Field

- Mean function – linear trend
- Covariance function – linear trend in variance
- Correlation structure – square exponential; correlation length $\lambda = 1\text{m}$
- Updating – single snow depth measure

(multiple measures improve prediction)

VoI for Decision Support: Local Failure

http://www.adn.com/sites/default/files/styles/ad_slideshow_wide/public/legacy/2012/02/edoog.So.7.jpg?itok=kkYvLv91



Effects of management actions

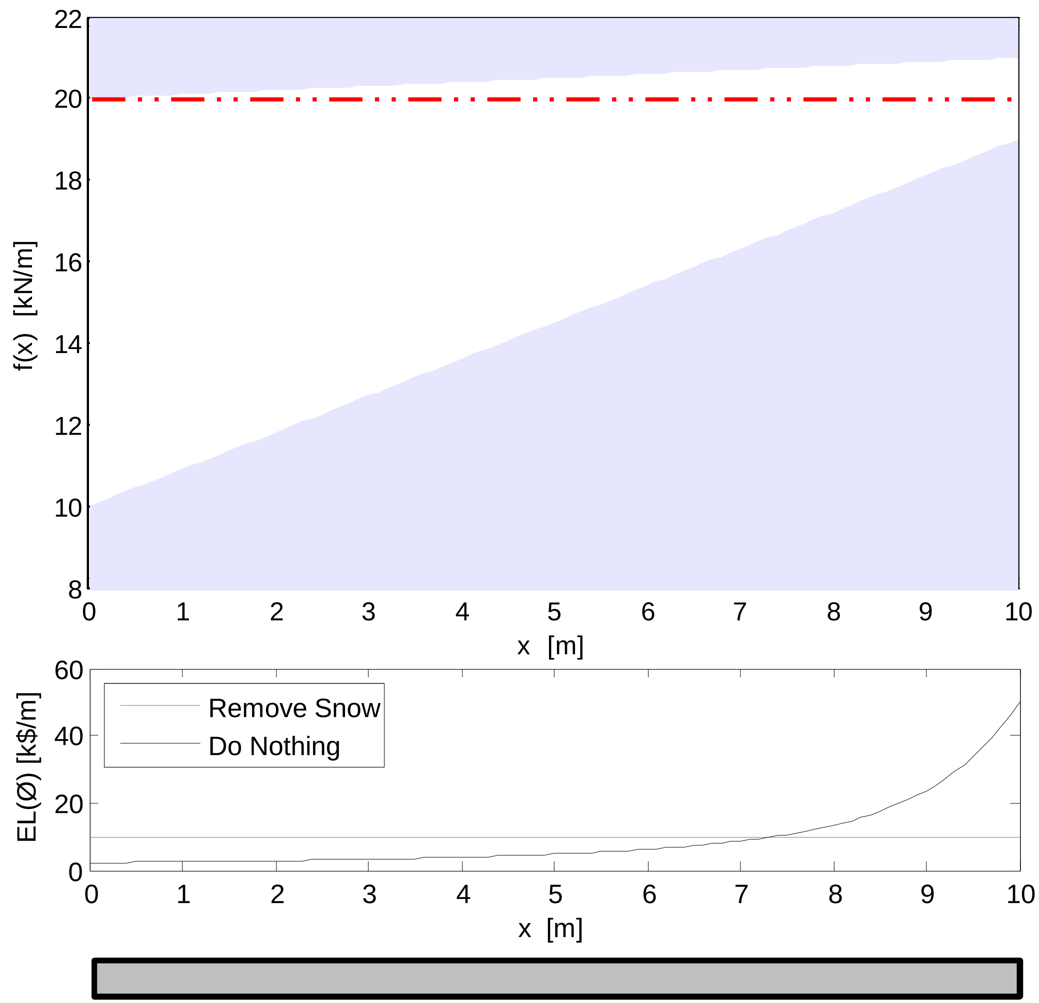
- Several actions possible
- Management loss depends on selected actions a and random field f: $\mathrm{L}(f,a)$

VoI for Decision Support: Local Failure

$l(f,a)$ (f)

(a)

	S	F	
N	0	C_F	100K$/m
R	C_R	C_R	10K$/m



actions

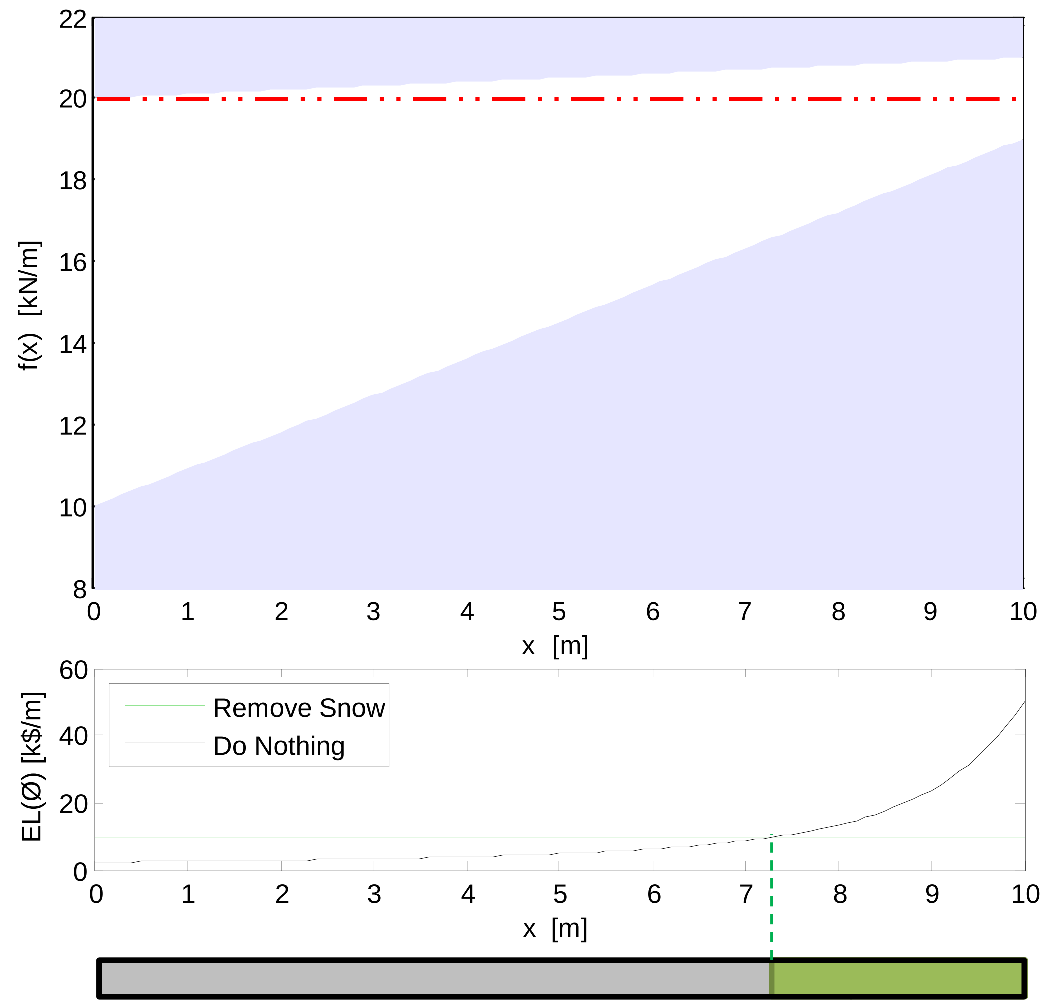
Value of Information

- Measures the expected reduction in management costs with additional data

VoI for Decision Support: Local Failure

$l(f,a)$

(a) \ (f)	S	F	
N	0	C_F	100K$/m
R	C_R	C_R	10K$/m



prior optimal actions

Value of Information

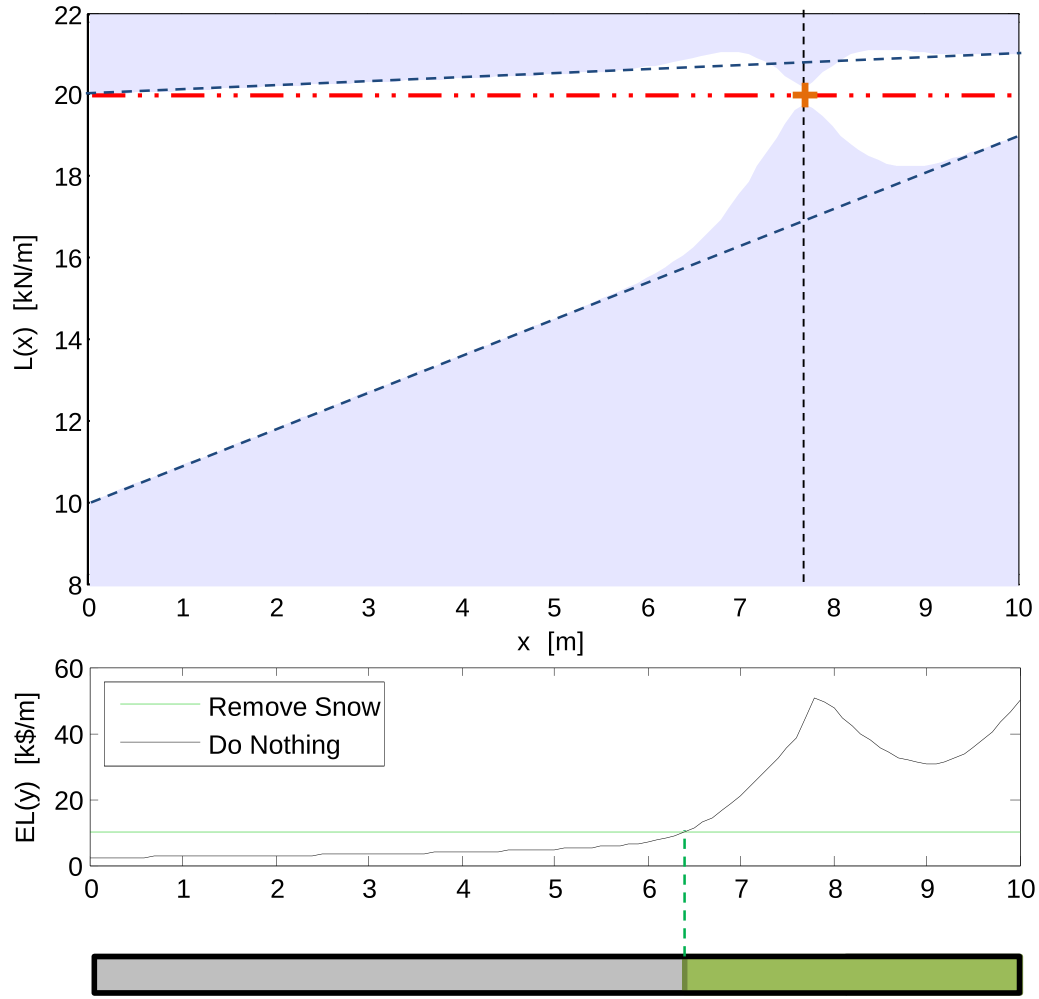
- Measures the expected reduction in management costs with additional data

VoI for Decision Support: Local Failure

$l(f, a)$ (f)

(a)

	S	F	
N	0	C_F	100K$/m
R	C_R	C_R	10K$/m



posterior optimal actions

Value of Information

- Measurement near "border" of prior decision regions
- Sensor placements support posterior decision-making

VoI for Decision Support: Local Failure

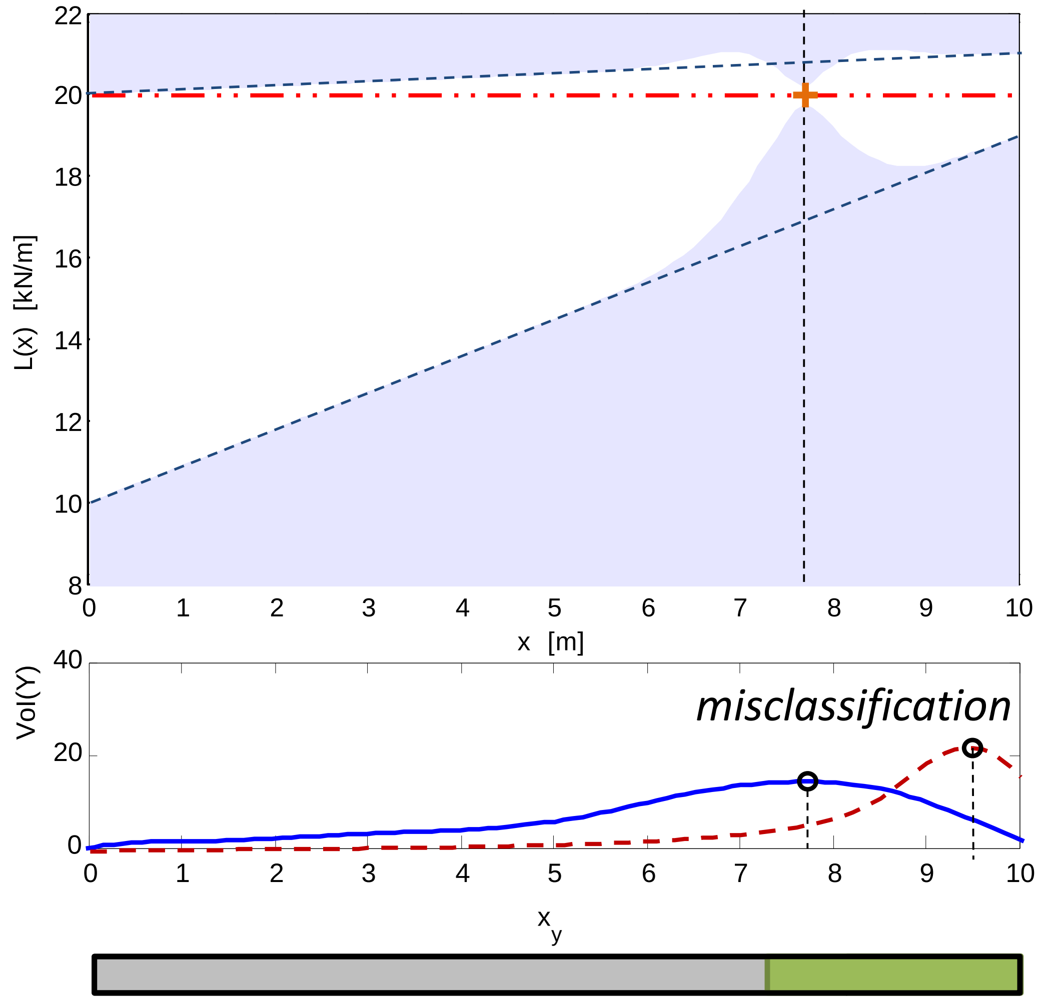
$l(f,a)$

(a) \ (f)	S	F	
N	0	C_F	100K$/m
R	C_R	C_R	10K$/m

misclassification: $C_F = 2C_R$

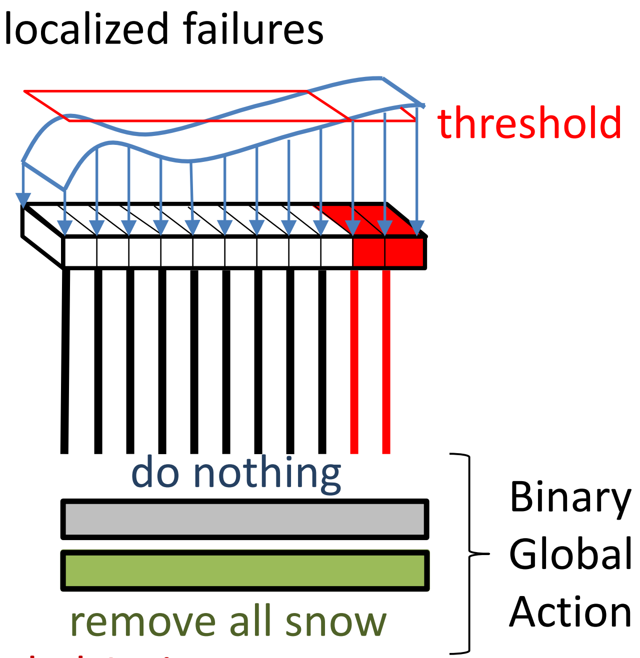
Value of Information

- Measurement near "border" of prior decision regions
- Sensor placements support posterior decision-making



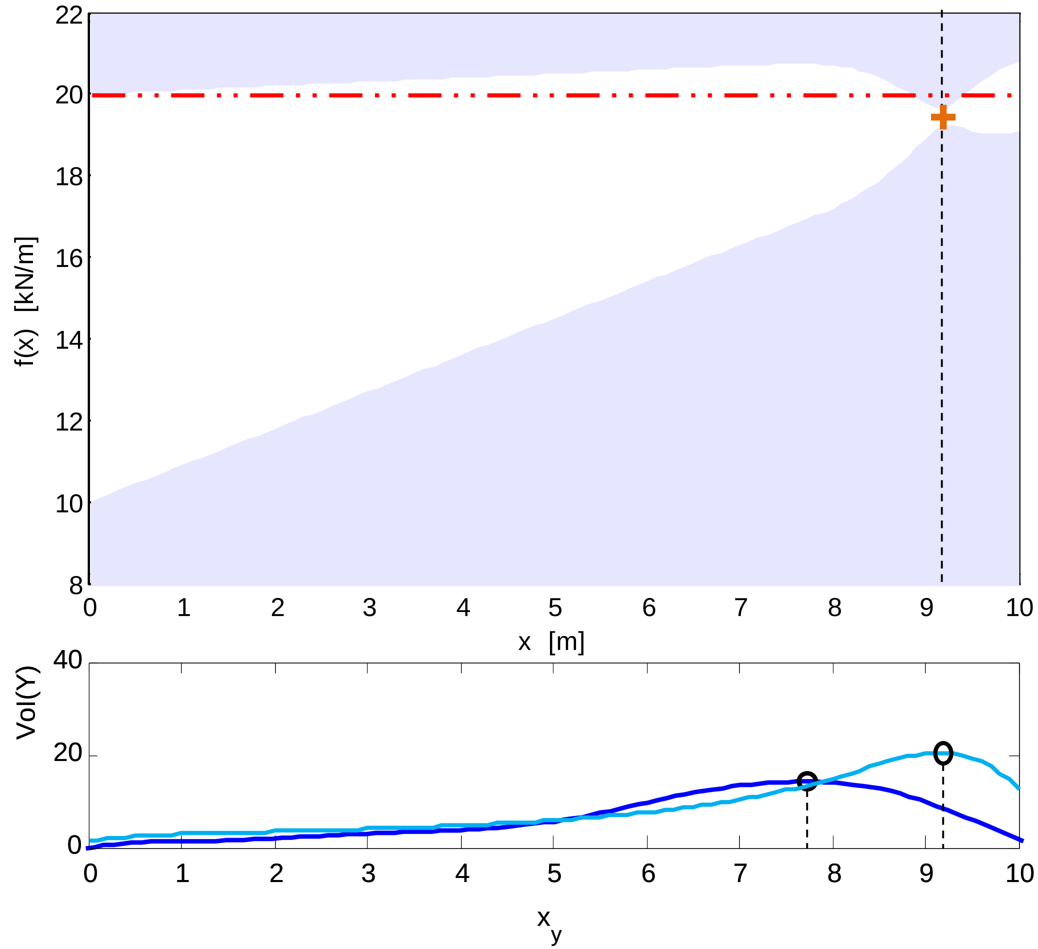
prior optimal actions

VoI for Decision Support: Global Actions



Global Actions

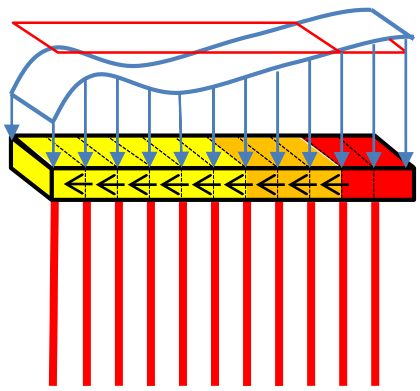
- Couple decision-making at system level
- No longer a cumulative system
- Increase computational demand



observe higher-risk areas – if no danger is observed, “do nothing” is likely best

VoI for Decision Support: Series System

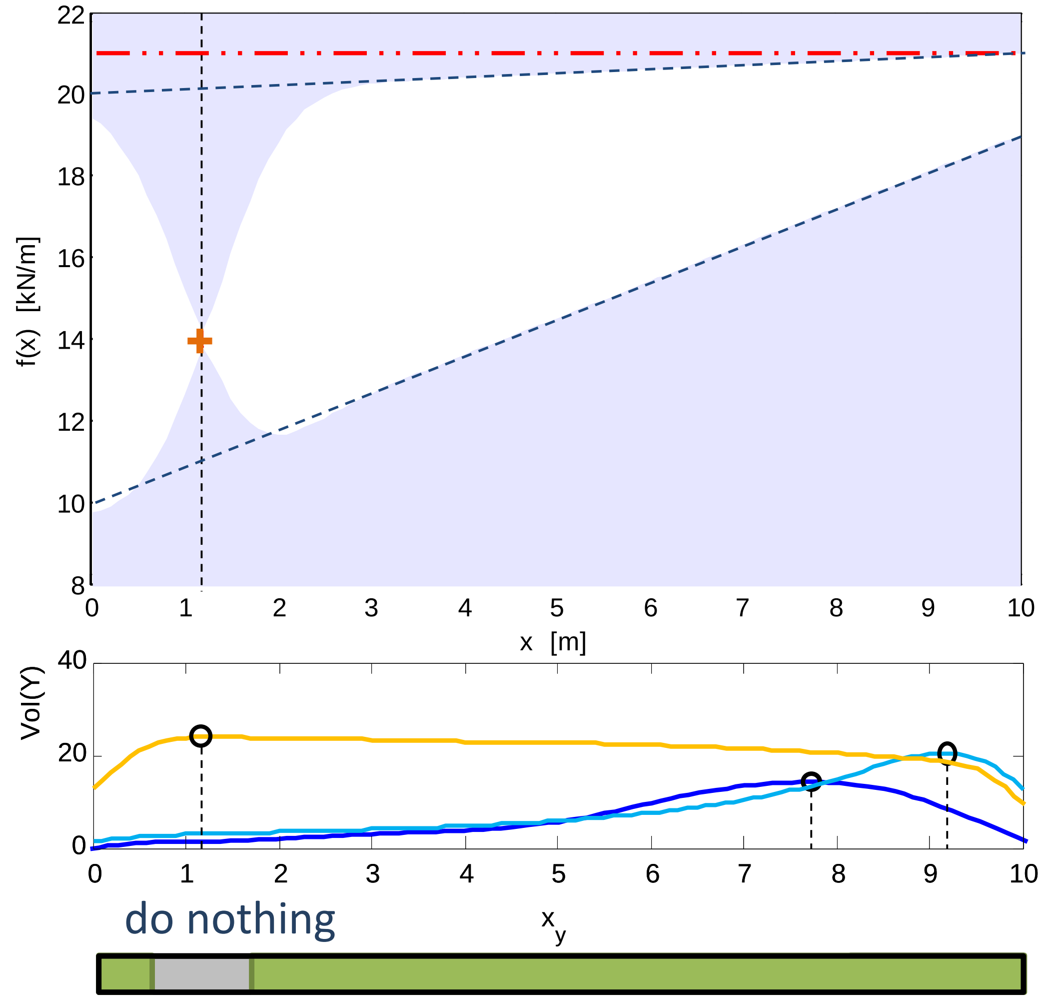
progressive collapse



$$\max_{i \in \{1,\dots,n\}} f(x_i) > T \rightarrow \color{red}{Failure}$$

System Topology Effects

- Couple decision-making at system level
- No longer a cumulative system
- Increase computational demand



posterior optimal actions remove snow

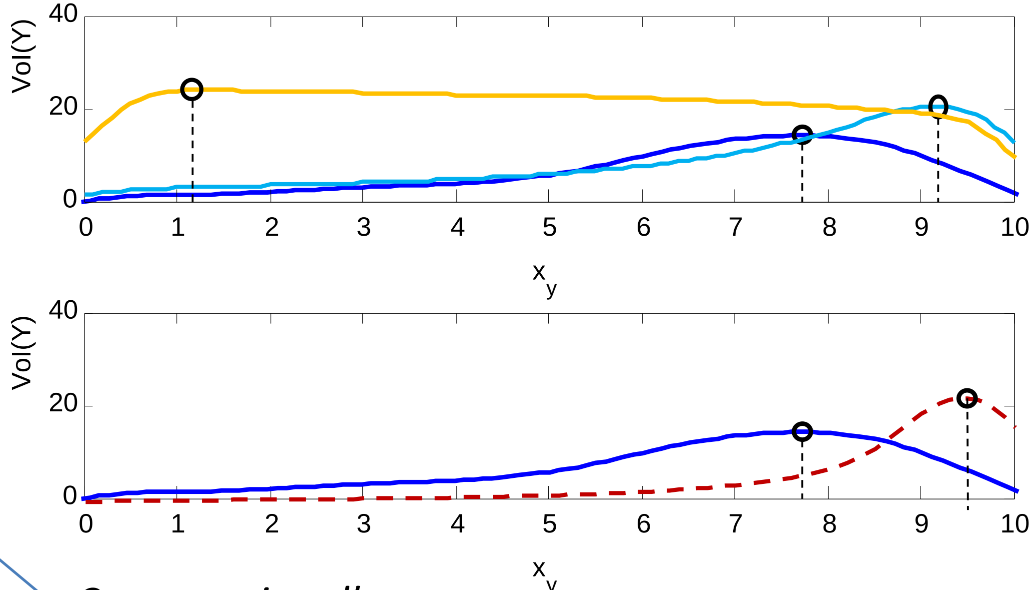
VoI: Sensor Placement with Gaussian Models

Efficient evaluation of VoI Is possible in cumulative systems with Gaussian variables.

Optimal placement strongly depends on loss and actions.

VoI computational complexity strongly depends on the decision making problem, specially when the number of sensors to be placed is high.

Complexity grows significantly when system functionality is included.



Malings, C., Pozzi, M. "Conditional entropy and value of information metrics for optimal sensing in infrastructure systems," Structural Safety 60:77-90. (Elsevier) doi:10.1016/j.strusafe.2015.10.003 (2016).

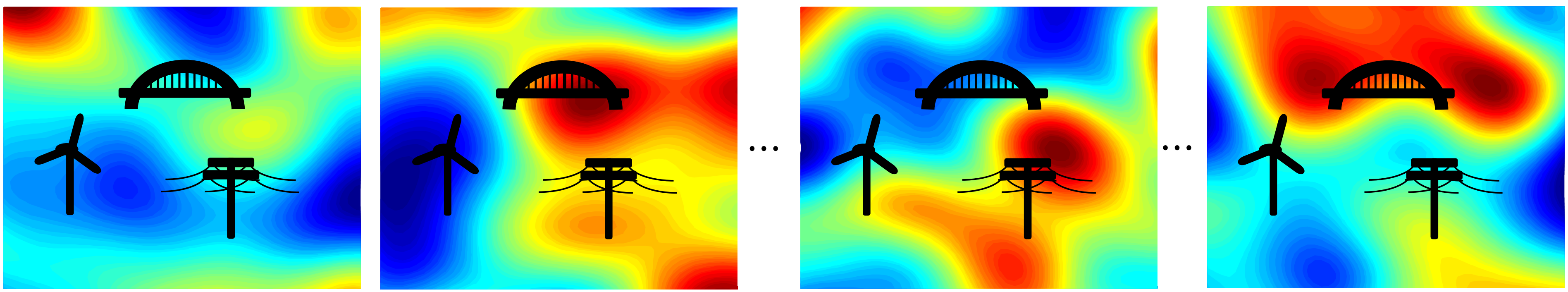
Malings, C., Pozzi, M. "Value of Information for Spatially Distributed Systems: application to sensor placement," submitted to Reliability Engineering & System Safety.

Computationally efficient formulation for local failures.

Spatio-Temporal Random Fields

Random variables distributed over space and time can be modeled as a spatio-temporal random field, e.g. a Gaussian spatio-temporal random field.

Similarities in variables are modeled via spatial and temporal covariance functions, e.g. assuming separability.

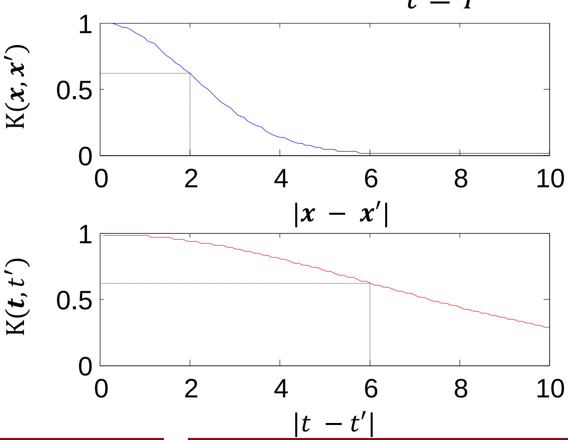


$t = 0$ $t = 1$ t $t = T$

$$f(\mathbf{x}, t) \sim \mathcal{GP}\big(\mathrm{M}(\mathbf{x}, t), \mathrm{K}(\mathbf{x}, t, \mathbf{x}', t')\big)$$

e.g. separable spatio-temporal covariance

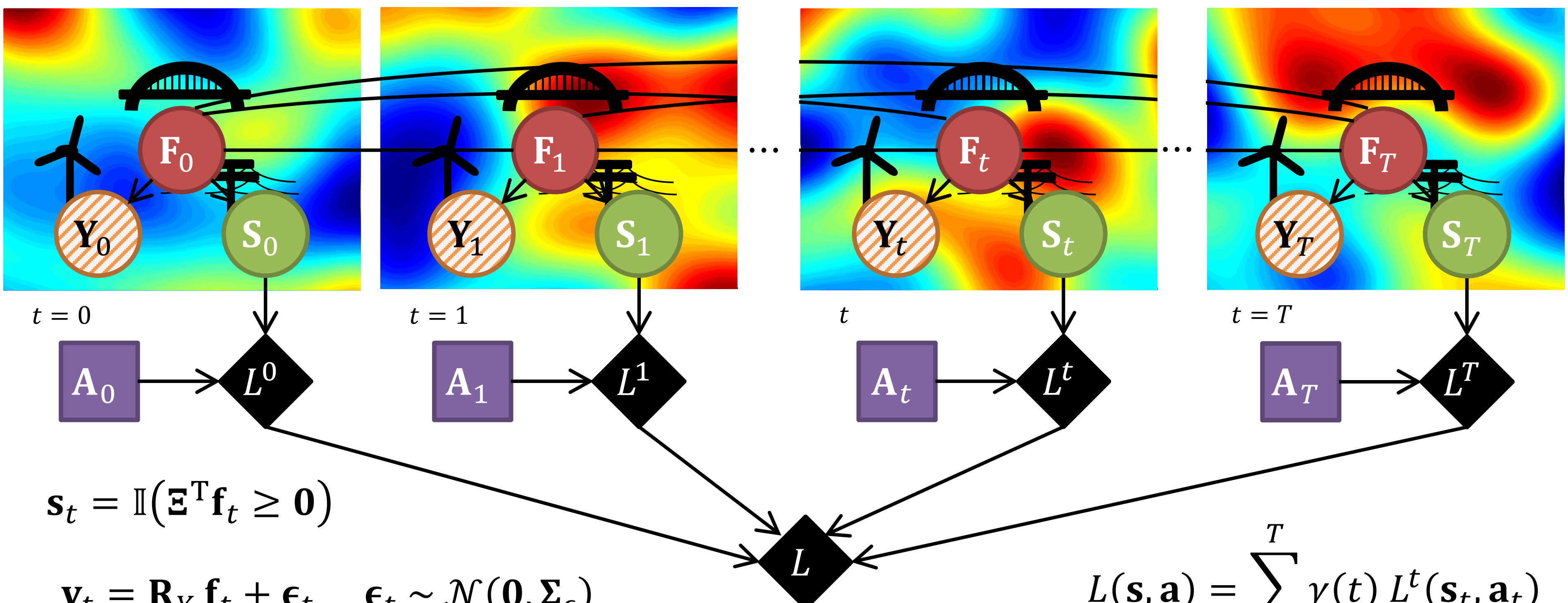
$$\mathrm{K}(\mathbf{x}, t, \mathbf{x}', t') = \mathrm{K}_X(\mathbf{x}, \mathbf{x}')\mathrm{K}_T(t, t')$$



Spatio-Temporal Random Field Models

The spatio-temporal random field model is combined with observation, state, action, and loss models for the entire time-horizon of the management problem.

This model captures information gathering and decision-making over time as the system changes in response to the evolving spatio-temporal random field.



$$\mathbf{s}_t = \mathbb{I}\left(\boldsymbol{\Xi}^{\mathrm{T}}\mathbf{f}_t \geq \mathbf{0}\right)$$

$$\mathbf{y}_t = \mathbf{R}_{Y_t}\mathbf{f}_t + \boldsymbol{\epsilon}_t \qquad \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$$

$$L(\mathbf{s}, \mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t, \mathbf{a}_t)$$

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

spatial case

$$\mathbb{E}L(Y) = \mathbb{E}_Y\left[\min_{\mathbf{a}} \mathbb{E}_{S|\mathbf{y}} L(\mathbf{s},\mathbf{a})\right]$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} L\big(\mathbf{s}, \mathbf{a}^*(\mathbf{y})\big)$$

$$\mathbf{a}^*(\mathbf{y}) = \{\mathbf{a}_1^*(\mathbf{y}_1^-), \dots, \mathbf{a}_t^*(\mathbf{y}_t^-), \dots, \mathbf{a}_T^*(\mathbf{y}_T^-)\}$$

$$\mathbf{y}_t^- = \{\mathbf{y}_1, \dots, \mathbf{y}_{t-1}\}$$

$for\ t = 0,1,\dots,T:$

$$\{\mathbf{a}_t^*(\mathbf{y}_t^-), \dots\} = \underset{\{\mathbf{a}_t,\dots,\mathbf{a}_T\}}{\operatorname{argmin}} \mathbb{E}_{S|\mathbf{y}_t^-} L(\mathbf{s}, \{\mathbf{a}_1^*(\mathbf{y}_1^-), \dots, \mathbf{a}_t, \dots, \mathbf{a}_T\})$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s},\mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} L\big(\mathbf{s},\mathbf{a}^*(\mathbf{y})\big)$$

$$\mathbf{a}^*(\mathbf{y}) = \{\mathbf{a}_1^*(\mathbf{y}_1^-),\dots,\mathbf{a}_t^*(\mathbf{y}_t^-),\dots,\mathbf{a}_T^*(\mathbf{y}_T^-)\}$$

$for\ t = 0,1,\dots,T:$

$$\{\mathbf{a}_t^*(\mathbf{y}_t^-),\dots\} = \operatorname*{argmin}_{\{\mathbf{a}_t,\dots,\mathbf{a}_T\}} \mathbb{E}_{S|\mathbf{y}_t^-} L(\mathbf{s},\{\mathbf{a}_1^*(\mathbf{y}_1^-),\dots,\mathbf{a}_t,\dots,\mathbf{a}_T\})$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s},\mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} L\big(\mathbf{s},\mathbf{a}^*(\mathbf{y})\big)$$

$$\mathbf{a}^*(\mathbf{y}) = \{\mathbf{a}_1^*(\mathbf{y}_1^-),\dots,\mathbf{a}_t^*(\mathbf{y}_t^-),\dots,\mathbf{a}_T^*(\mathbf{y}_T^-)\}$$

$$for\ t = 0,1,\dots,T:$$

$$\{\mathbf{a}_t^*(\mathbf{y}_t^-),\dots\} = \underset{\{\mathbf{a}_t,\dots,\mathbf{a}_T\}}{\operatorname{argmin}} \mathbb{E}_{S|\mathbf{y}_t^-}\left[\textstyle\sum_{\tau=1}^{T} \gamma(\tau)\, L^\tau(\mathbf{s}_\tau,\mathbf{a}_\tau)\right]$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s},\mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} L\big(\mathbf{s},\mathbf{a}^*(\mathbf{y})\big)$$

$$\mathbf{a}^*(\mathbf{y}) = \{\mathbf{a}_1^*(\mathbf{y}_1^-),\dots,\mathbf{a}_t^*(\mathbf{y}_t^-),\dots,\mathbf{a}_T^*(\mathbf{y}_T^-)\}$$

constant w.r.t. $\mathbf{a}_t,\dots,\mathbf{a}_T$

$for\ t = 0,1,\dots,T:$

$$\{\mathbf{a}_t^*(\mathbf{y}_t^-),\dots\} = \underset{\{\mathbf{a}_t,\dots,\mathbf{a}_T\}}{\operatorname{argmin}} \mathbb{E}_{S|\mathbf{y}_t^-}\left[\begin{array}{c}\cancel{\sum_{\tau=1}^{t-1}\gamma(\tau)\, L^\tau\big(\mathbf{s}_\tau,\mathbf{a}_\tau^*(\mathbf{y}_\tau^-)\big)} + \\ \gamma(t)\, L^\tau(\mathbf{s}_t,\mathbf{a}_t) + \\ \cancel{\sum_{\tau=t+1}^{T}\gamma(\tau)\, L^\tau(\mathbf{s}_\tau,\mathbf{a}_\tau)}\end{array}\right]$$

no impact on $\mathbf{a}_t$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s},\mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} L\big(\mathbf{s},\mathbf{a}^*(\mathbf{y})\big)$$

$$\mathbf{a}^*(\mathbf{y}) = \{\mathbf{a}_1^*(\mathbf{y}_1^-),\dots,\mathbf{a}_t^*(\mathbf{y}_t^-),\dots,\mathbf{a}_T^*(\mathbf{y}_T^-)\}$$

$$for\ t = 0,1,\dots,T:$$

constant

$$\mathbf{a}_t^*(\mathbf{y}_t^-) = \operatorname*{argmin}_{\mathbf{a}_t} \mathbb{E}_{S_t|\mathbf{y}_t^-}\, \cancel{\gamma(t)}\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s}, \mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t, \mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \mathbb{E}_{S,Y} \sum_{t=1}^{T} \gamma(t)\, L^t\big(\mathbf{s}_t, \mathbf{a}_t^*(\mathbf{y}_t^-)\big)$$

$$for\ t = 0,1, \dots, T:$$

$$\mathbf{a}_t^*(\mathbf{y}_t^-) = \underset{\mathbf{a}_t}{\operatorname{argmin}}\, \mathbb{E}_{S_t|\mathbf{y}_t^-}\, L^t(\mathbf{s}_t, \mathbf{a}_t)$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s},\mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \sum_{t=1}^{T} \gamma(t)\,\mathbb{E}_{Y_t^-}\,\mathbb{E}_{S_t|\mathbf{y}_t^-}\, L^t\big(\mathbf{s}_t, \mathbf{a}_t^*(\mathbf{y}_t^-)\big)$$

$for\ t = 0,1,\dots,T\ :$

$$\mathbf{a}_t^*(\mathbf{y}_t^-) = \operatorname*{argmin}_{\mathbf{a}_t} \mathbb{E}_{S_t|\mathbf{y}_t^-}\, L^t(\mathbf{s}_t,\mathbf{a}_t)$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

Value of Information in Spatio-Temporal Systems

Key difference from spatial case: spatio-temporal VoI can only be evaluated using information collected before a decision is made.

As with a spatially decomposable loss function, a temporally decomposable loss function allows for efficient evaluation of temporal VoI.

temporal decomposability

$$L(\mathbf{s}, \mathbf{a}) = \sum_{t=1}^{T} \gamma(t)\, L^t(\mathbf{s}_t, \mathbf{a}_t)$$

spatio-temporal case

$$\mathbb{E}L(Y) = \sum_{t=1}^{T} \gamma(t) \mathbb{E}_{Y_t^-} \min_{\mathbf{a}_t} \mathbb{E}_{S_t|\mathbf{y}_t^-} L^t(\mathbf{s}_t, \mathbf{a}_t)$$

temporal decomposability $\rightarrow$ VoI can be evaluated as a (temporally) local quantity

$$for\ t = 0,1, \dots, T:$$

$$\mathbf{a}_t^*(\mathbf{y}_t^-) = \operatorname*{argmin}_{\mathbf{a}_t} \mathbb{E}_{S_t|\mathbf{y}_t^-} L^t(\mathbf{s}_t, \mathbf{a}_t)$$

Malings, C., Pozzi, M. *Value of Information in Spatio-Temporal Systems: sensor placement and scheduling.* Reliability Engineering and System Safety, In Review.

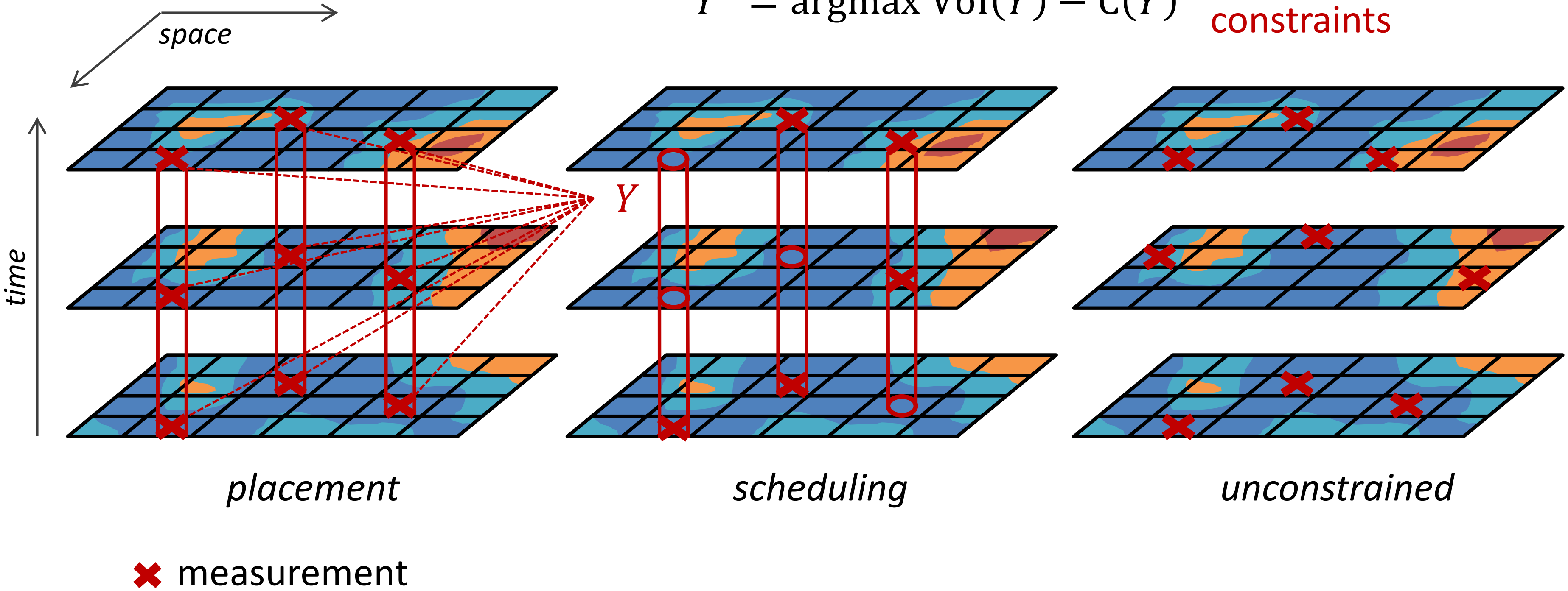
Sensor Placement and Scheduling

The problems of sensor placement and sensor scheduling represent different constraints on the general problem of selecting the optimal subset locations and times to measure in the spatio-temporal random field problem.

Optimal Sensing Objective:

$$Y^* = \text{argmax VoI}(Y) - \text{C}(Y)$$

cost function encodes constraints

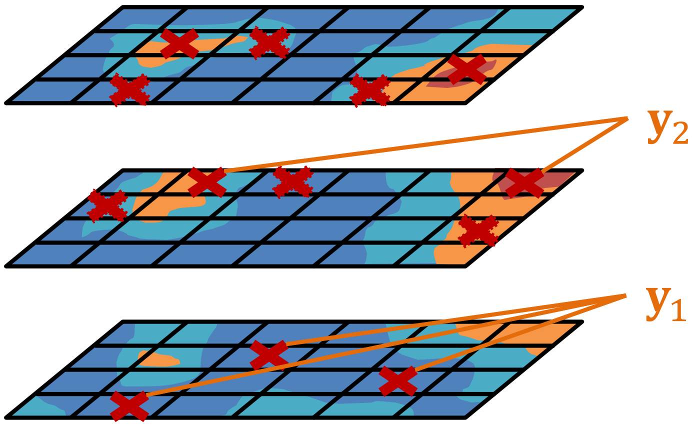


Sensor Placement and Scheduling: Online v. Offline

Offline: measure selection is performed before any information is gathered.

Online: gathered information supports future measure selection.

The online case can be treated as iterated offline placement



$$Y^* = \text{argmax}\ \text{VoI}(Y) - \text{C}(Y)$$

1. Observe Y_1
2. Update model using $\mathbf{y}_1$
3. Re-evaluate Y^* (keeping Y_1 fixed)
4. Observe Y_2

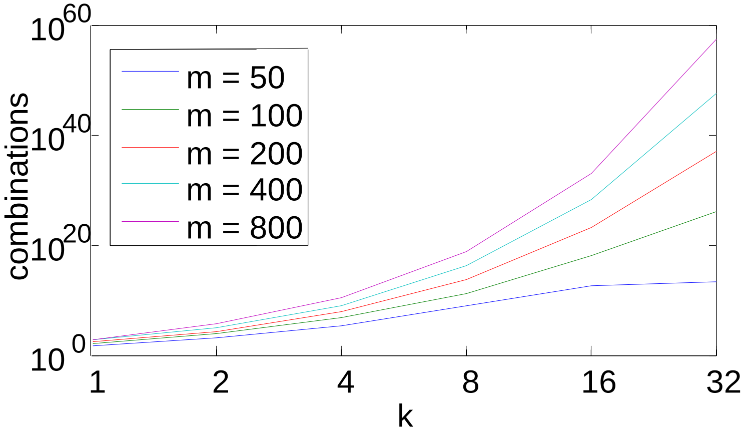
⋮

Sensor Placement and Scheduling: Greedy Optimization

Selecting an optimal sensing set is a problem in combinatorial optimization.

With many possible observations, exact optimization is intractable.

An approximate method, such as greedy optimization, can be applied.



$$Y^* = \text{argmax VoI}(Y) - \text{C}(Y)$$

candidate measures: $m = |Tn|$

number of sensors: $k = |Y|$

combinations: $\binom{m}{k}$

greedy algorithm:

Initialize $Y = \emptyset$

Repeat: $y^* = \text{argmax VoI}(y \cup Y) - \text{C}(y \cup Y)$

$Y \leftarrow y^* \cup Y$

Until the maximum net VoI is achieved.

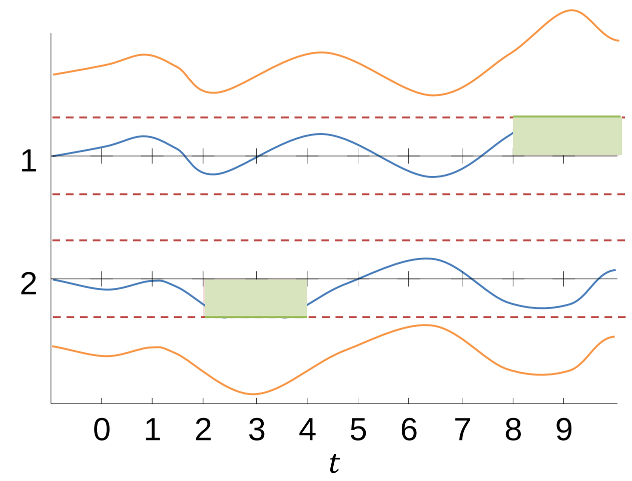
# of computations:

mk instead of $\binom{m}{k}$

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions



Example

- Two components (described with two variables) evolving in time
- Failure if absolute value exceeded
- Mitigation actions possible
- Biased measurements possible
- Same cost per measurement

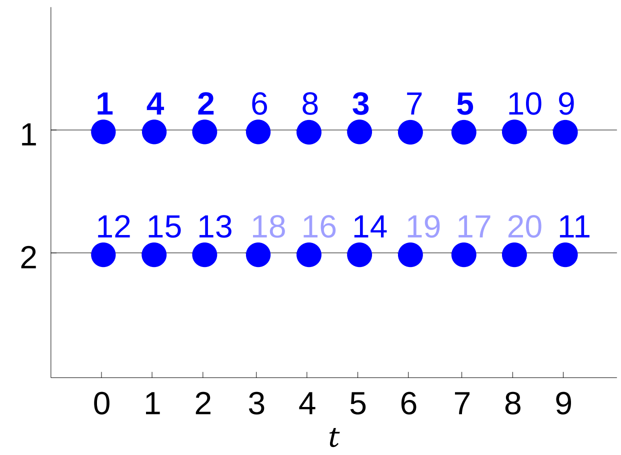
Krause, A., 2008. *Optimizing Sensing: Theory and Applications.* Pittsburgh, PA.

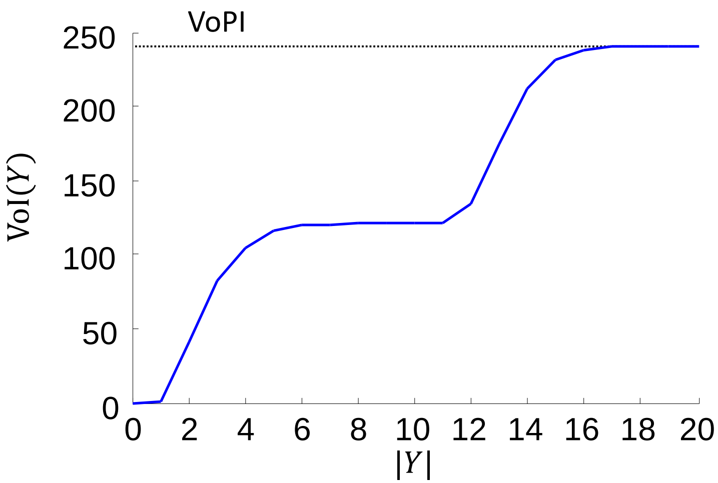
Krause, A., and Guestrin, C., 2009. "Optimal Value of Information in Graphical Models." *Journal of Artificial Intelligence Research*, 35: 557-591.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions





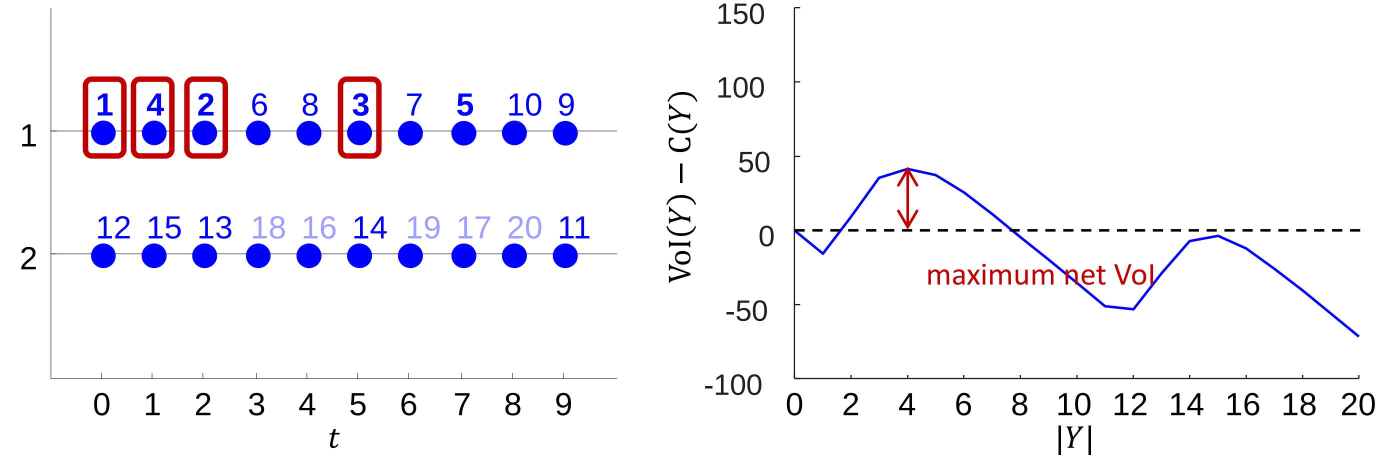
- Forward Greedy Optimization

Krause, A., and Guestrin, C., 2009. "Optimal Value of Information in Graphical Models." *Journal of Artificial Intelligence Research*, 35: 557-591.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions



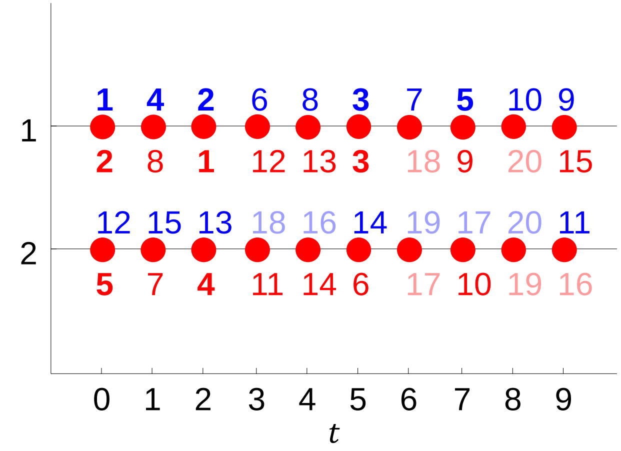
- Forward Greedy Optimization

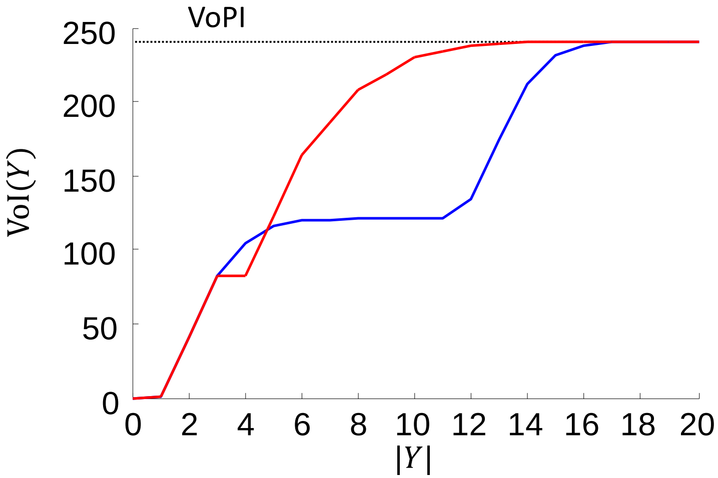
Krause, A., and Guestrin, C., 2009. "Optimal Value of Information in Graphical Models." *Journal of Artificial Intelligence Research*, 35: 557-591.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions





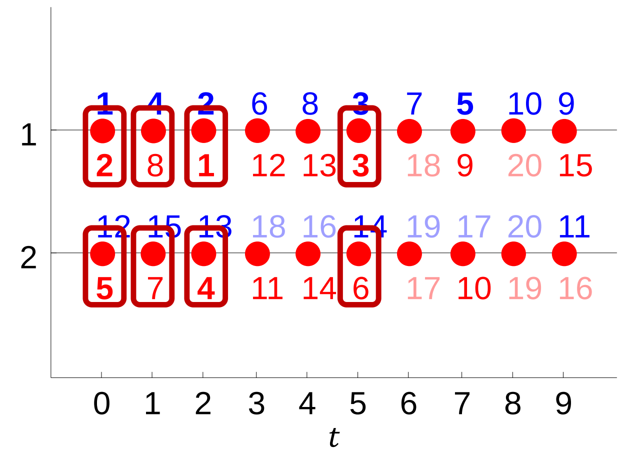
- Forward Greedy Optimization
- Reverse Greedy Optimization

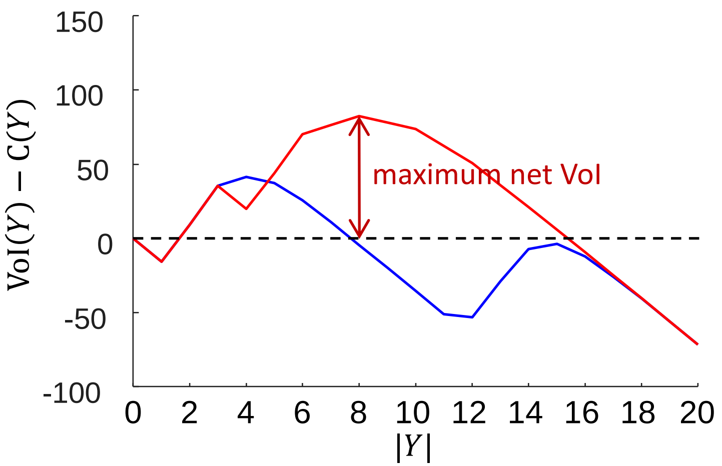
Papadimitriou, C., 2004. "Optimal sensor placement methodology for parametric identification of structural systems." *Journal of Sound and Vibration,* 278: 923–947.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions





- Forward Greedy Optimization
- Reverse Greedy Optimization

Papadimitriou, C., 2004. "Optimal sensor placement methodology for parametric identification of structural systems." *Journal of Sound and Vibration,* 278: 923–947.

Shortcomings of Greedy Optimization

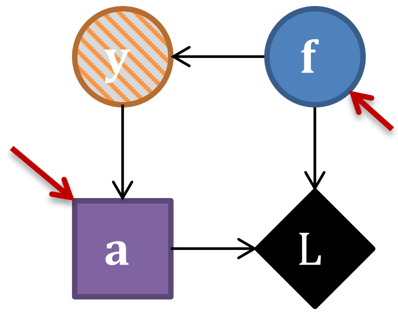
Heuristic Solution: Use Alternative Loss Functions

Value of Information

$\mathrm{VoI}(Y) = \mathbb{E}L(\emptyset) - \mathbb{E}L(Y)$

emphasizes loss reduction

non-submodular



Prediction Error

$\Delta\mathrm{Err}(Y) = \mathrm{tr}(\boldsymbol{\Sigma}_f) - \mathrm{tr}(\boldsymbol{\Sigma}_{f|Y})$

emphasizes uncertainty reduction

submodular

Proposed Heuristic Greedy Approach

begin by optimizing $\mathrm{VoI}(Y)$ →

when VoI growth stalls, optimize $\Delta\mathrm{Err}(Y)$

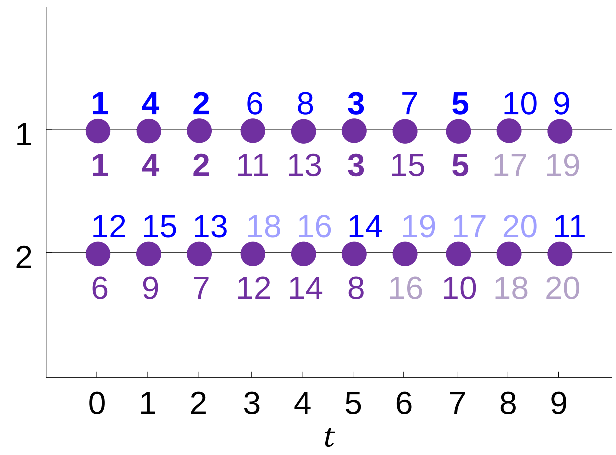
← return to optimizing $\mathrm{VoI}(Y)$

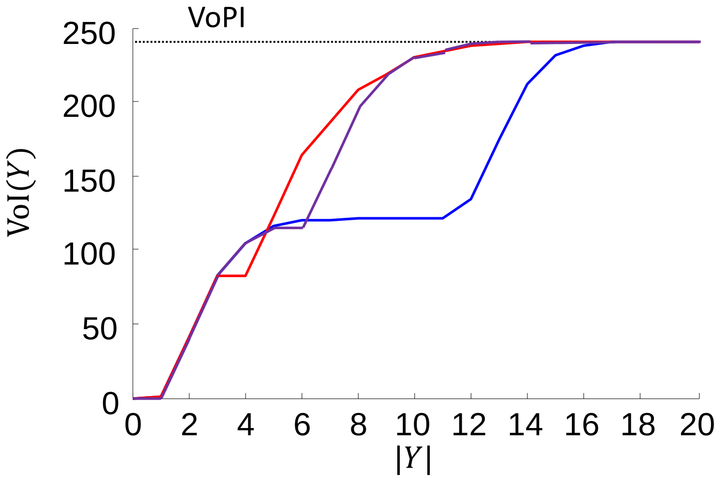
Malings, C., Pozzi, M. Submodularity Issues in Value-of-Information-based Sensor Placement. *Reliability Engineering and System Safety*, manuscript under review.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions





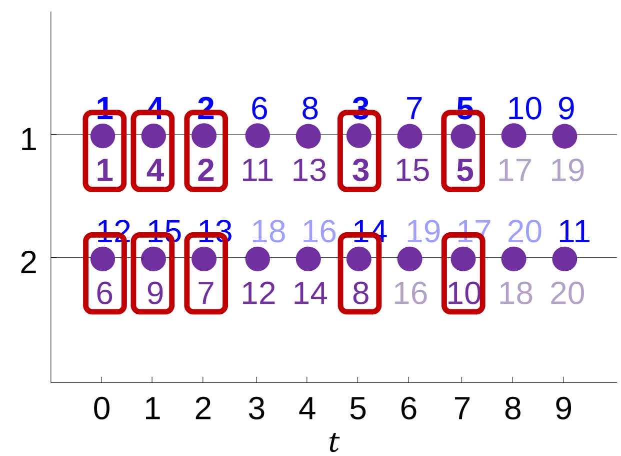
- Forward Greedy Optimization
- Reverse Greedy Optimization
- Optimization with Heuristics (conditional entropy)

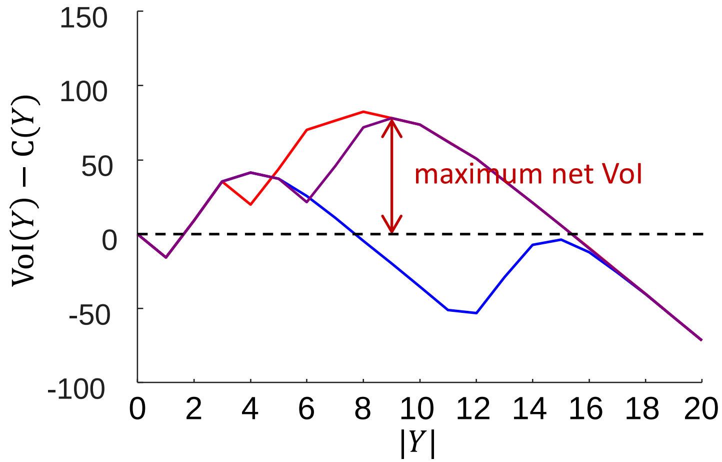
Malings, C., Pozzi, M. Submodularity Issues in Value-of-Information-based Sensor Placement. *Reliability Engineering and System Safety*, manuscript under review.

Shortcomings of Greedy Optimization

The Value of Information metric lacks the property of Submodularity

Greedy optimization approaches can lead to highly suboptimal solutions





- Forward Greedy Optimization
- Reverse Greedy Optimization
- Optimization with Heuristics (conditional entropy)

Malings, C., Pozzi, M. Submodularity Issues in Value-of-Information-based Sensor Placement. *Reliability Engineering and System Safety,* manuscript under review.

Summary and Conclusions

Value of information represents an important tool for optimizing measurement selection and inspection planning to support decision-making and system management.

However, the computational cost of its implementation depends on the size of the problem, in terms of number of possible states, actions, observations, and future steps.

Assumptions about the system functionality and the decision-making problem structure can be made which can reduce the computational cost.

Gaussian random fields are special class of system model for which Bayesian updating and computation is particularly easy.

For performing inference on a more general class, we need to adopt a numerical scheme, such as Markov Chain Monte Carlo.

To include model uncertainty into the VoI analysis is a challenging task still to be explored…

Additional References

Probabilistic Modeling

Rasmussen, C., Williams, C. (2006). Gaussian Processes for Machine Learning. The MIT Press. Downloadable from http://www.gaussianprocess.org/

Koller, D., Friedman, N. (2009). *Probabilistic Graphical Models Principles and Techniques*. the MIT Press.

Barber, B. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge UP. Downloadable from http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage

Value of Information

Raiffa, H., Schlaifer, R. (1961). *Applied statistical decision theory.* John Wiley & Sons, Inc., New York, NY.

Pozzi M., Der Kiureghian A., (2011). "Assessing the value of information for long-term structural health monitoring", *SPIE Proceedings* vol. 7984, , (Tribikram Kundu, Ed.). SPIE Conference, Health Monitoring of Structural and Biological Systems. San Diego, CA.

Sensor Placement

Krause, A. (2008). *Optimizing Sensing: Theory and Applications.* Carnegie Mellon University PhD Dissertation, Pittsburgh, PA.

Malings, C., Pozzi, M. (2016). "Conditional Entropy and Value of Information Metrics for Optimal Sensing in Infrastructure Systems", *Structural Safety*, vol. 60, pp. 77-90.

Malings, C., Pozzi, M. (2016). "Value of Information for Spatially Distributed Systems: application to sensor placement", *Reliability Engineering & System Safety*, vol. 154, pp. 219-233.

Inspection Scheduling

Memarzadeh, M., Pozzi, M., Kolter, J.Z. (2015) "Optimal planning and learning in uncertain environments for the management of wind farm," *ASCE Journal of Computing in Civil Engineering*, vol. 29, issue 5.

Memarzadeh, M., Pozzi, M. (2015). "Integrated inspection scheduling and maintenance planning for infrastructure systems," *Computer-Aided Civil and Infrastructure Engineering* (Wiley) DOI: 10.1111/mice.12178.

Memarzadeh, M., Pozzi, M. (2016). "Value of Information in Sequential Decision Making: component inspection, permanent monitoring and system-level scheduling," *Reliability Engineering & System Safety*, vol. 154, pp. 137-151.

value of information in
spatio-temporal random fields

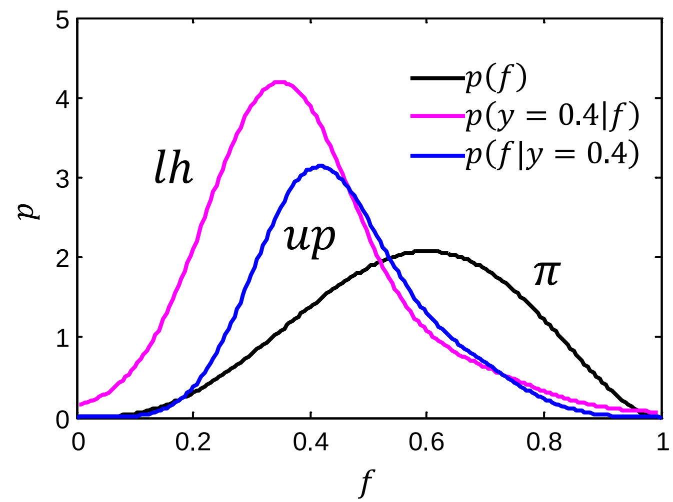
thanks for your attention!

Backup Materials

Terminology of Bayesian Inference

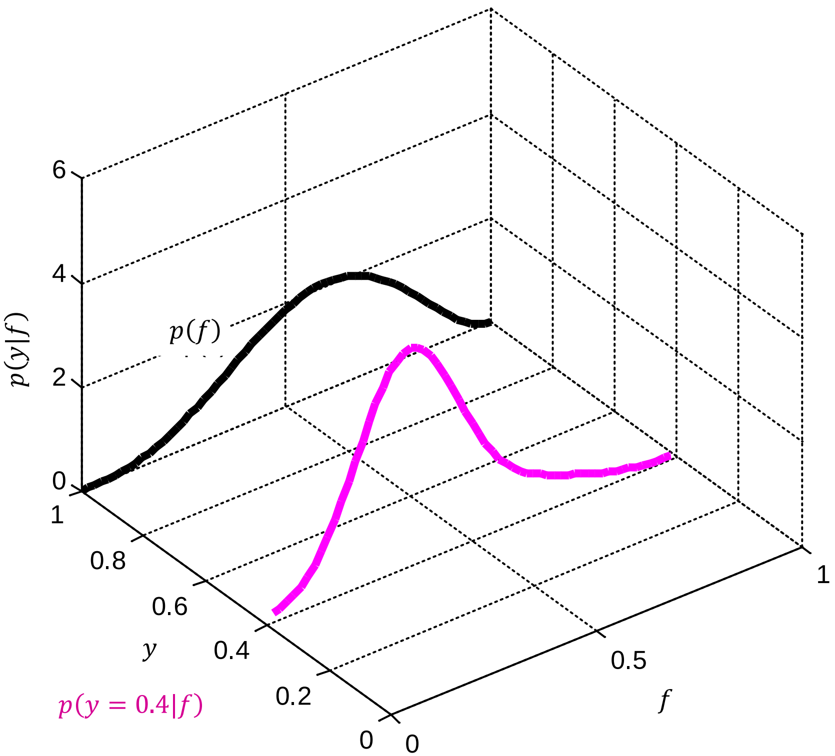
prior distribution	$\pi(f) \triangleq p(f)$	$\pi(f) \geq 0;\ \int_{-\infty}^{\infty} \pi(f) df = 1$
likelihood function, for observation $y = \tilde{y}$	$lh(f) \triangleq p(y = \tilde{y} \mid f)$	$lh(f) \geq 0;\ \int_{-\infty}^{\infty} lh(f) df$ *not necessary one*
posterior distribution	$up(f) \triangleq p(f \mid y = \tilde{y})$	$up(f) \geq 0;\ \int_{-\infty}^{\infty} up(f) df = 1$

Bayes' formula $\boxed{up(f) \propto \pi(f) lh(f)}$



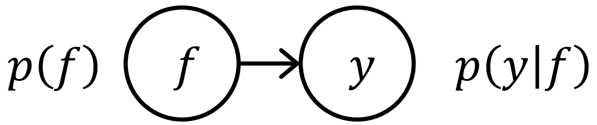
Processing Observation y

Assumptions on $p(f)$ and $p(y|f)$ are sufficient to define joint probability:



chain rule

$p(f, y) = p(f)\, p(y|f)$



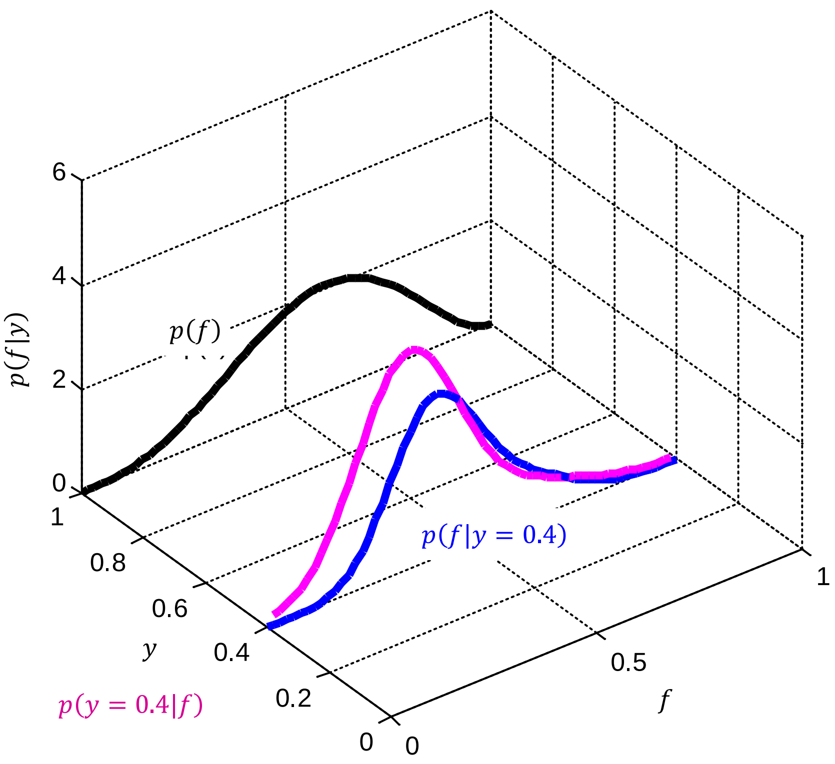
Likelihood $p(y|f)$ defines observation y for any value of f.

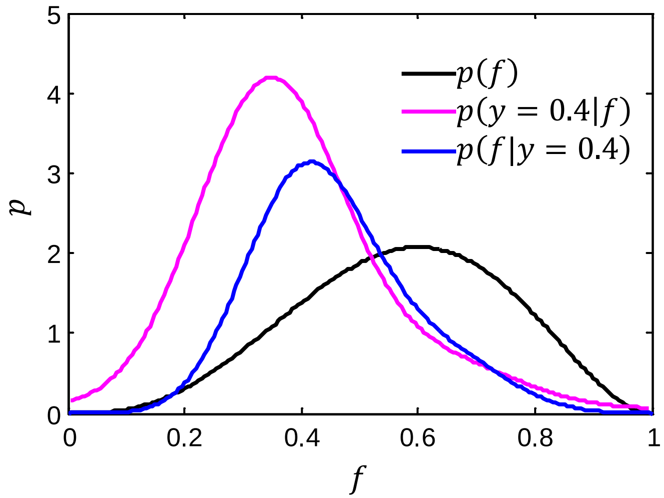
Suppose to observe $y = \tilde{y}$:

Only univariate (likelihood) function $p(y = \tilde{y}|f)$ is relevant.

Bayes' Formula

posterior probability: $$p(f|y=y') = \frac{p(f|y=y')p(f)}{\int_f p(f|y=y')p(f)df}$$





Multiplication of Normal Distributions

prior distribution $\pi(f) = \mathcal{N}\left(f; \mu_\pi, \sigma_\pi^2\right)$

likelihood function $lh(f) \propto \mathcal{N}\left(f; \mu_{\mathrm{LH}}, \sigma_{\mathrm{LH}}^2\right)$

posterior (updated) distribution $up(f) \propto \pi(f) lh(f) = \mathcal{N}\left(f; \mu_{\mathrm{up}}, \sigma_{\mathrm{up}}^2\right)$

the family of normal distributions is closed with respect to the product

parameters of the updated distribution:

$$\sigma_{\mathrm{up}}^2 = \left(\sigma_\pi^{-2} + \sigma_{\mathrm{LH}}^{-2}\right)^{-1} \qquad \frac{1}{\sigma_{\mathrm{up}}^2} = \frac{1}{\sigma_\pi^2} + \frac{1}{\sigma_{\mathrm{LH}}^2}$$

posterior unceratinty only depends on the variance of prior and LH func.

perfect information *perfect measure*

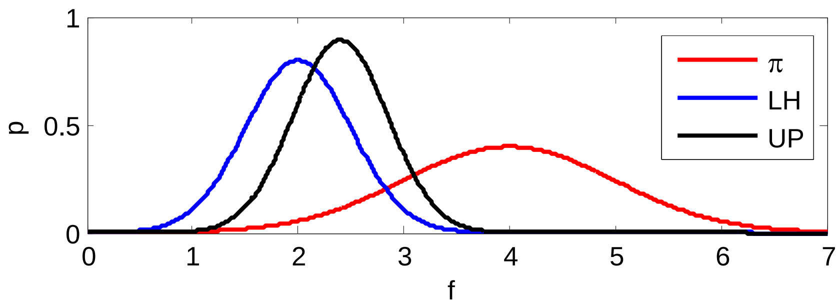
$$\sigma_\pi = 0 \text{ or } \sigma_{\mathrm{LH}} = 0 \implies \sigma_{\mathrm{up}} = 0$$

independent measure: irrelevant

$$\sigma_{\mathrm{LH}} \to \infty \Longrightarrow \sigma_{\mathrm{up}} = \sigma_\pi$$

no prior information:
the posterior is equal to the LH

$$\sigma_\pi \to \infty \Longrightarrow \sigma_{\mathrm{up}} = \sigma_{\mathrm{LH}}$$



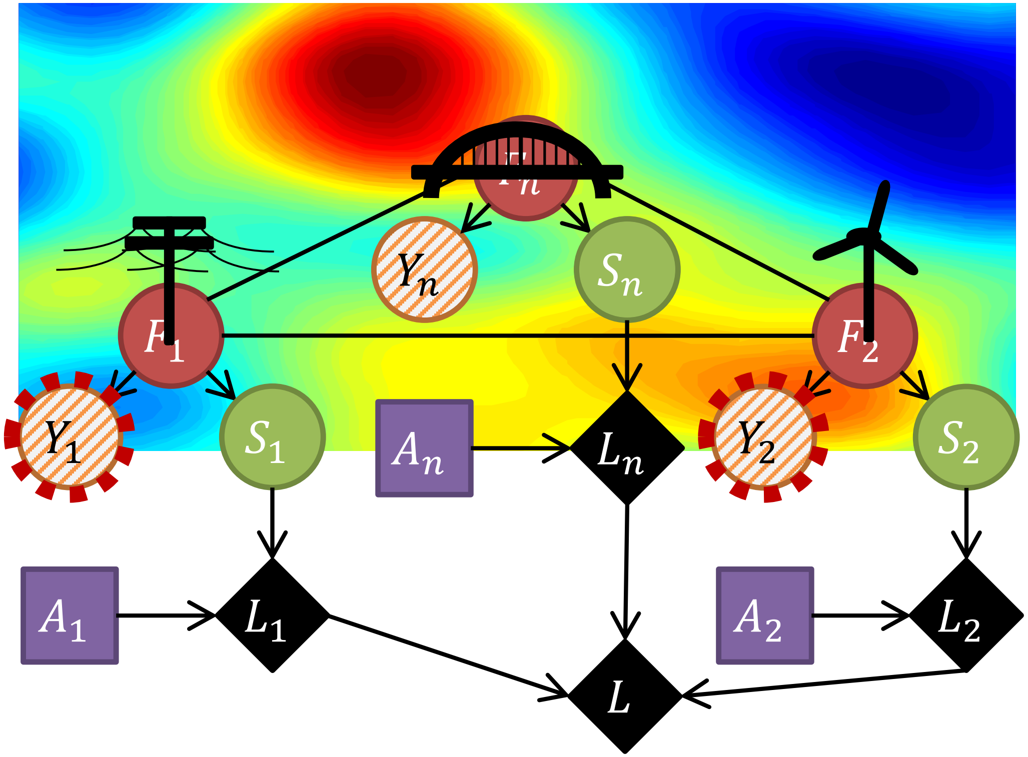
general case: reduction of uncertainty

$$\sigma_{\mathrm{up}} \leq \min(\sigma_\pi, \sigma_{\mathrm{LH}})$$

Value of Information in Gaussian Random Fields

A model for system performance and decision-making combines the probabilistic random field model for the spatially distributed system (F) with models for observations (Y), components states (S), managing actions (A), and losses (L).

Under the cumulative system assumption and in Gaussian random fields, value of information can be efficiently evaluated to compare potential sensing schemes.



Observation Model:

$$\mathbf{y} = \mathbf{R}_Y \mathbf{f} + \boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$$

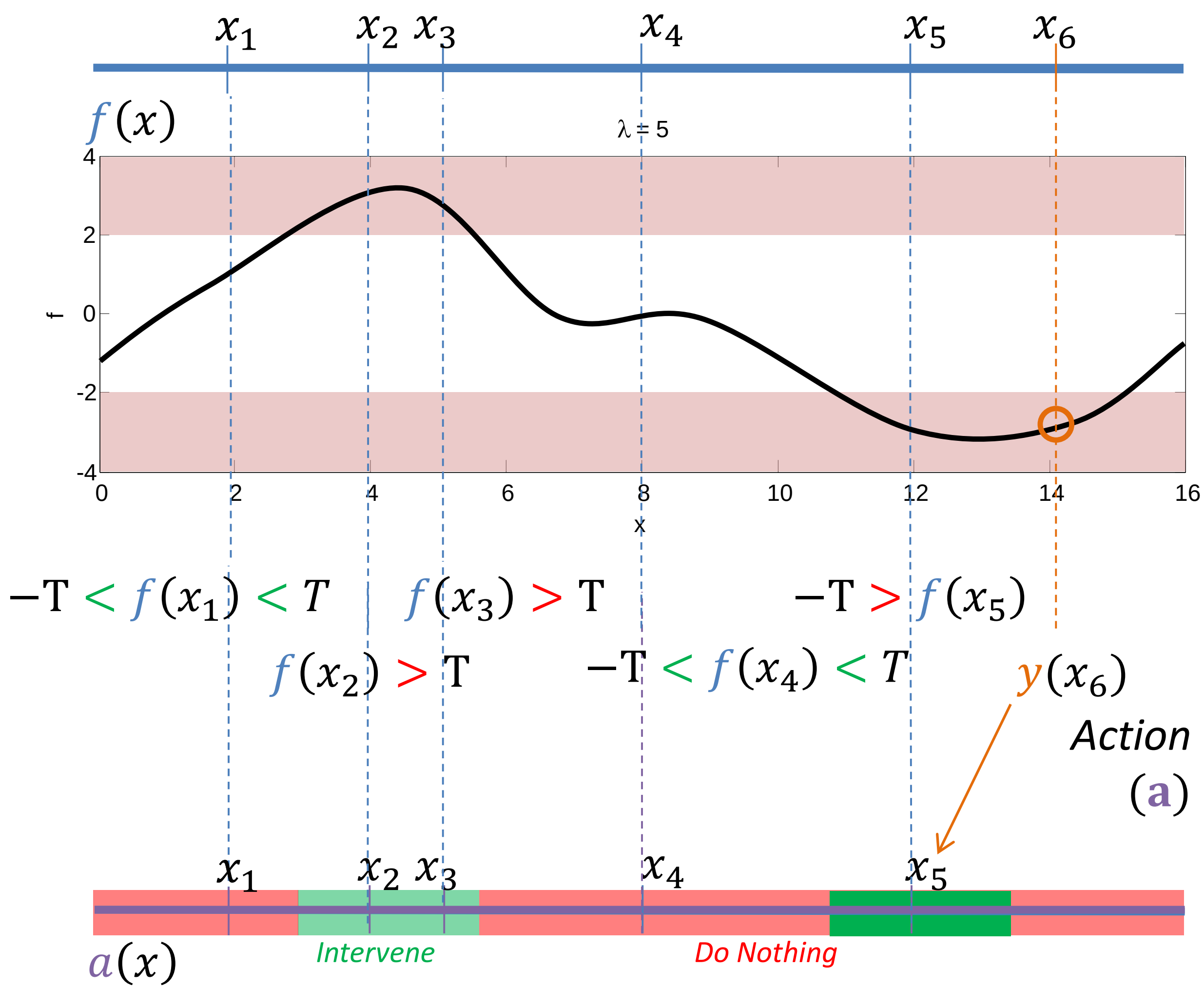
State Model:

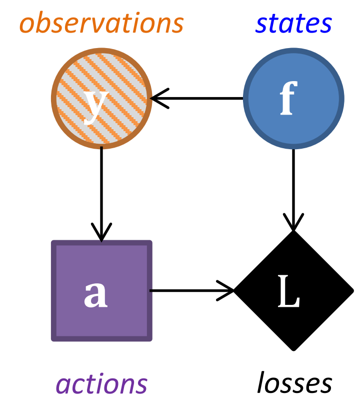
$$\mathbf{s} = \mathbb{I}\left(\boldsymbol{\Xi}^{\mathrm{T}} \mathbf{f} \geq \mathbf{0}\right)$$

Loss Model:

$$L(\mathbf{s}, \mathbf{a}) = \sum_{i=1}^{n} L_i(s_i, a_i)$$

Gaussian Random Field System Models





State (**f**)

Loss $L(\mathbf{f}, \mathbf{a})$	Safe	Unsafe
Action (a) Do Nothing	**0**	C_f
Intervene	C_I	C_I

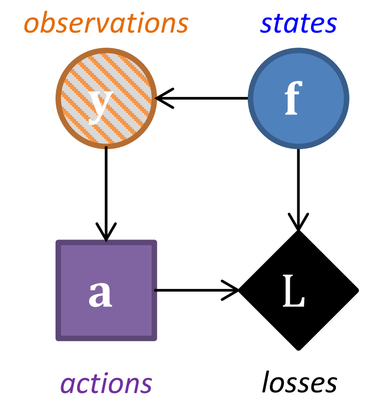
Value of Information

Quantification of loss reduction resulting from availability of additional information to support decision-making

$$\mathbb{E}\mathrm{L}(\emptyset) = \min_{A} \mathbb{E}_{F}\mathrm{L}(\mathbf{f}, \mathbf{a})$$

$$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_{Y} \min_{A} \mathbb{E}_{F|\mathbf{y}}\mathrm{L}(\mathbf{f}, \mathbf{a})$$

$$\mathrm{VoI}(Y) = \mathbb{E}\mathrm{L}(\emptyset) - \mathbb{E}\mathrm{L}(Y)$$



Value of Information at System-Level

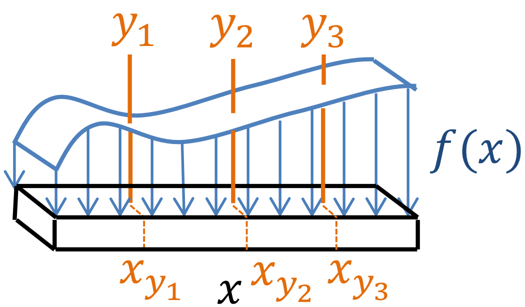
- Multiple components within the system can be in several states
- Multiple actions can be taken to manage components across the system
- Observations can provide information about single components, multiple components, or other variables related to component performance, i.e. information is shared across the system

Raiffa, H., and Schlaifer, R., 1961. *Applied Statistical Decision Theory.* Harvard University Press, Cambridge, MA.

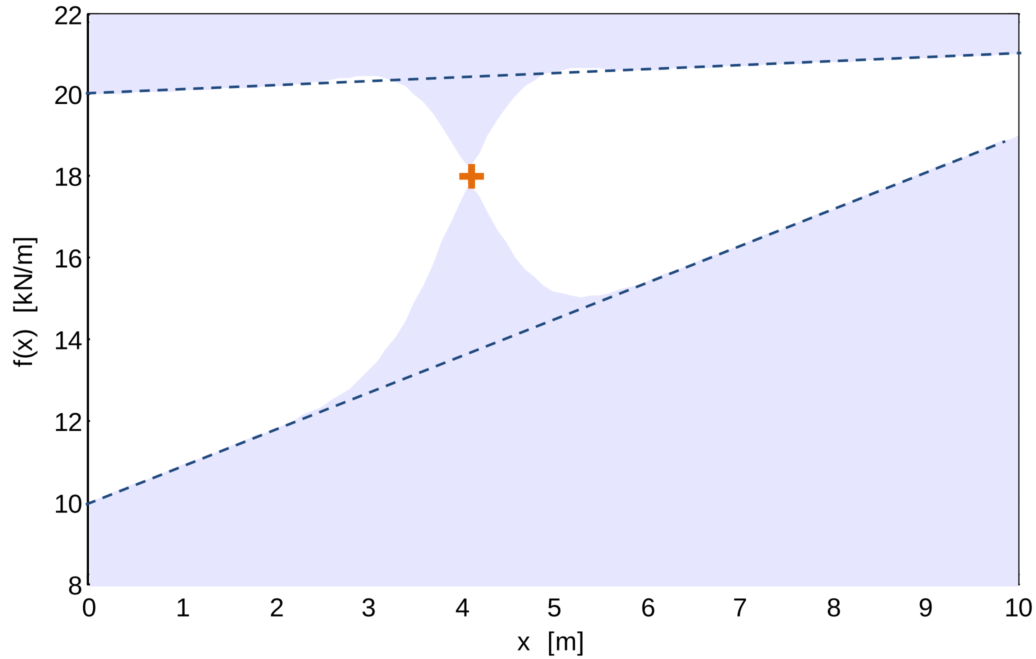
Sensor Placement

$$f(x) \sim \mathcal{GP}\big(\mu(x), K(x,x')\big)$$

$$\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{\mu}(X_Y), \boldsymbol{\Sigma}_{f(X_Y)} + \boldsymbol{\Sigma}_{\epsilon}\big)$$



$$\boldsymbol{f}|\boldsymbol{y} \sim \mathcal{N}\big(\boldsymbol{\mu}_{f|y}, \boldsymbol{\Sigma}_{f|Y}\big)$$



what is the best location to place a sensor?

- sensor reduces uncertainty
- they help in making decisions

- Optimal location depends on loss and actions!

VoI for Uncertainty Reduction: Random Field Prediction

$a(x) = \hat{f}(x)$

$l\big(f(x), a(x)\big) = [f(x) - a(x)]^2$

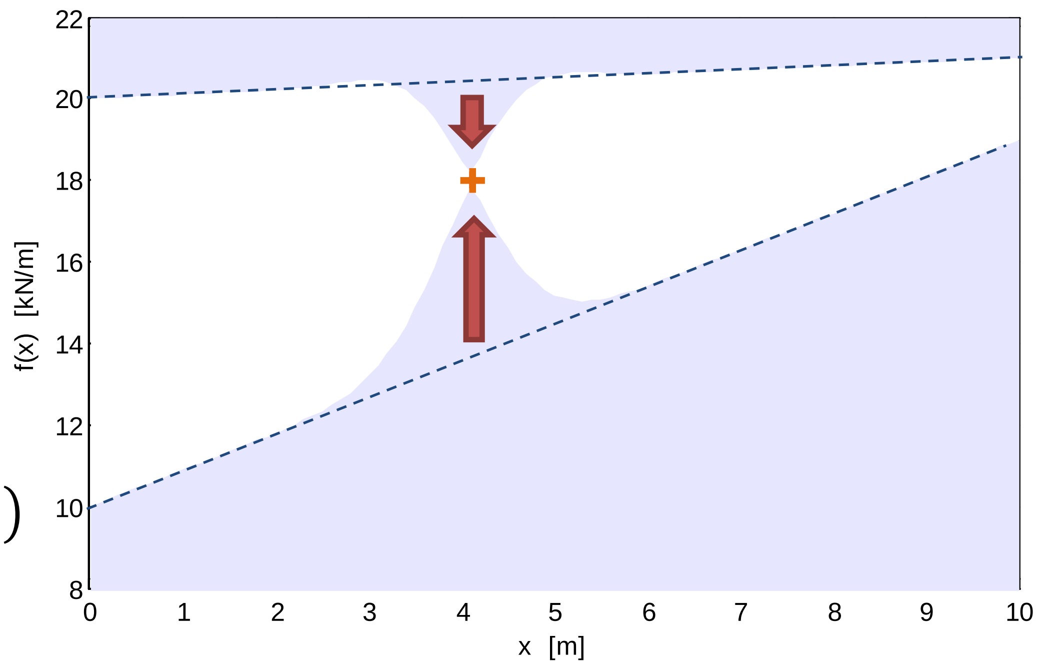
$\min_a \mathbb{E}_f l\big(f(x), a(x)\big) = \mathrm{Var}[f(x)]$

$\mathrm{L}(\mathbf{f}, \mathbf{a}) = \sum_{i=1}^{n} l\big(f(x_i), a(x_i)\big)$

$\mathbb{E}\mathrm{L}(\emptyset) = \min_A \mathbb{E}_F \mathrm{L}(\mathbf{f}, \mathbf{a}) = \mathrm{tr}\big(\boldsymbol{\Sigma}_f\big)$

$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \min_A \mathbb{E}_{F|\mathbf{y}} \mathrm{L}(\mathbf{f}, \mathbf{a}) = \mathrm{tr}\big(\boldsymbol{\Sigma}_{f|Y}\big)$

$\mathrm{VoI}(Y) = \mathrm{tr}\big(\boldsymbol{\Sigma}_f\big) - \mathrm{tr}\big(\boldsymbol{\Sigma}_{f|Y}\big)$



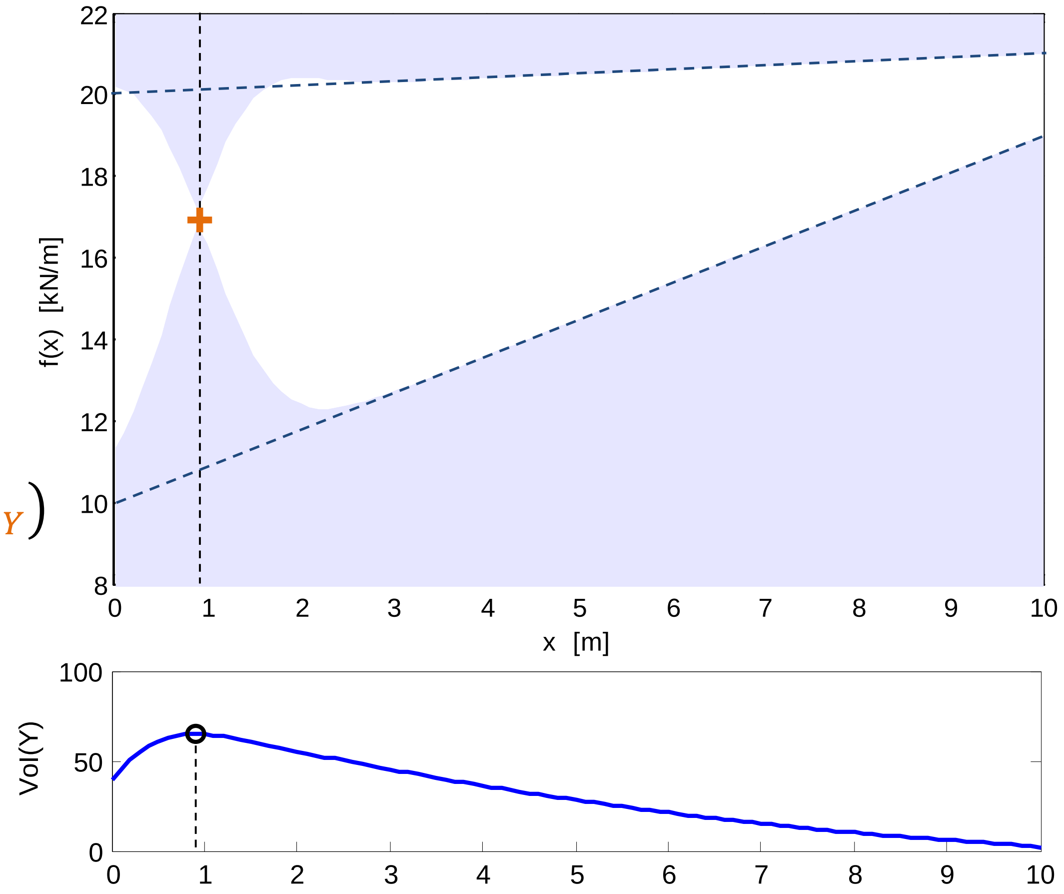
Uncertainty Reduction

- Uncertainty is reduced through sensing
- Place sensors to maximize uncertainty reduction
- Measure uncertainty using L-2 Norm error
- For Gaussian models, this metric has a closed-form expression

VoI for Uncertainty Reduction: Random Field Prediction

$a(x) = \hat{f}(x)$
$l(f(x), a(x)) = [f(x) - a(x)]^2$
$\min_a \mathbb{E}_f l(f(x), a(x)) = \mathrm{Var}[f(x)]$
$\mathrm{L}(\mathbf{f}, \mathbf{a}) = \sum_{i=1}^{n} l(f(x_i), a(x_i))$

$\mathbb{E}\mathrm{L}(\emptyset) = \min_A \mathbb{E}_F \mathrm{L}(\mathbf{f}, \mathbf{a}) = \mathrm{tr}(\boldsymbol{\Sigma}_f)$
$\mathbb{E}\mathrm{L}(Y) = \mathbb{E}_Y \min_A \mathbb{E}_{F|\mathbf{y}} \mathrm{L}(\mathbf{f}, \mathbf{a}) = \mathrm{tr}(\boldsymbol{\Sigma}_{f|Y})$
$\mathrm{VoI}(Y) = \mathrm{tr}(\boldsymbol{\Sigma}_f) - \mathrm{tr}(\boldsymbol{\Sigma}_{f|Y})$



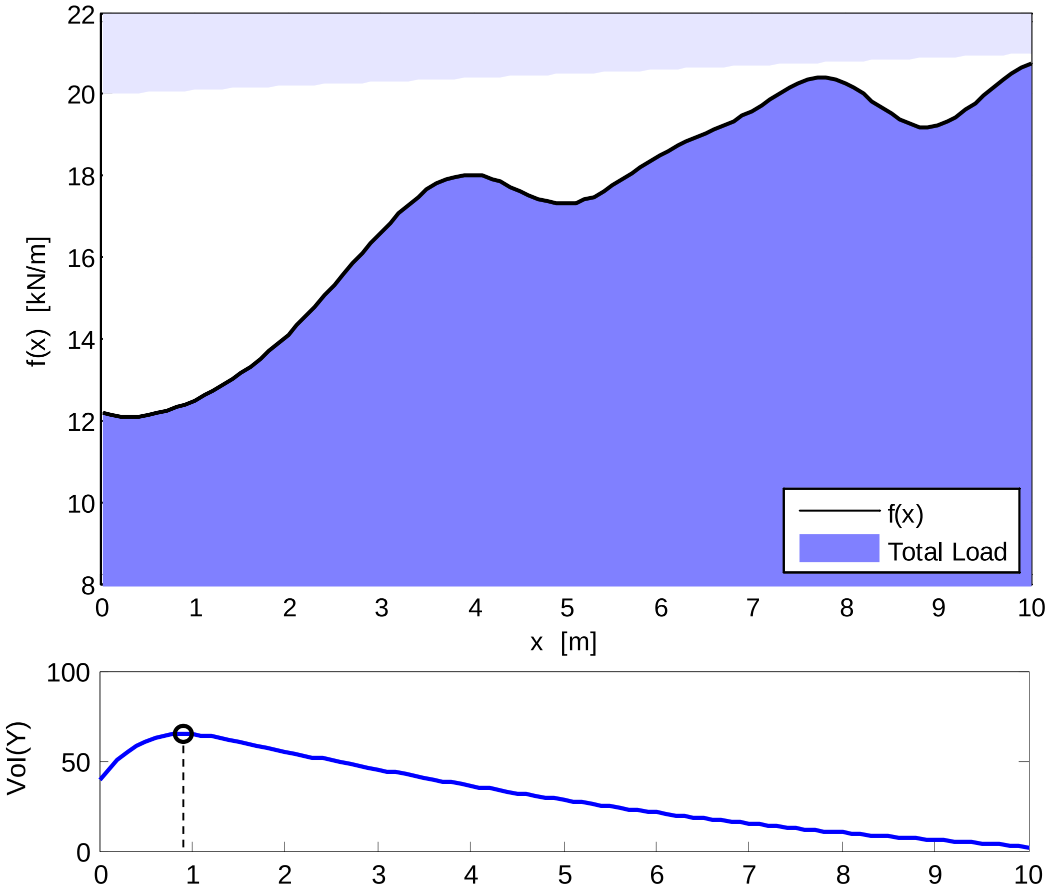
Uncertainty Reduction

- Optimal location on left side
- Targets most uncertain areas
- Uncertainty in adjacent areas is also reduced

VoI for Uncertainty Reduction: Global Load

$$F = \sum_{i=1,\ldots,n} f(x_i) = \mathbf{1}^{\mathrm{T}} \boldsymbol{f}$$

$$\boldsymbol{f} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f\right)$$



Global loading

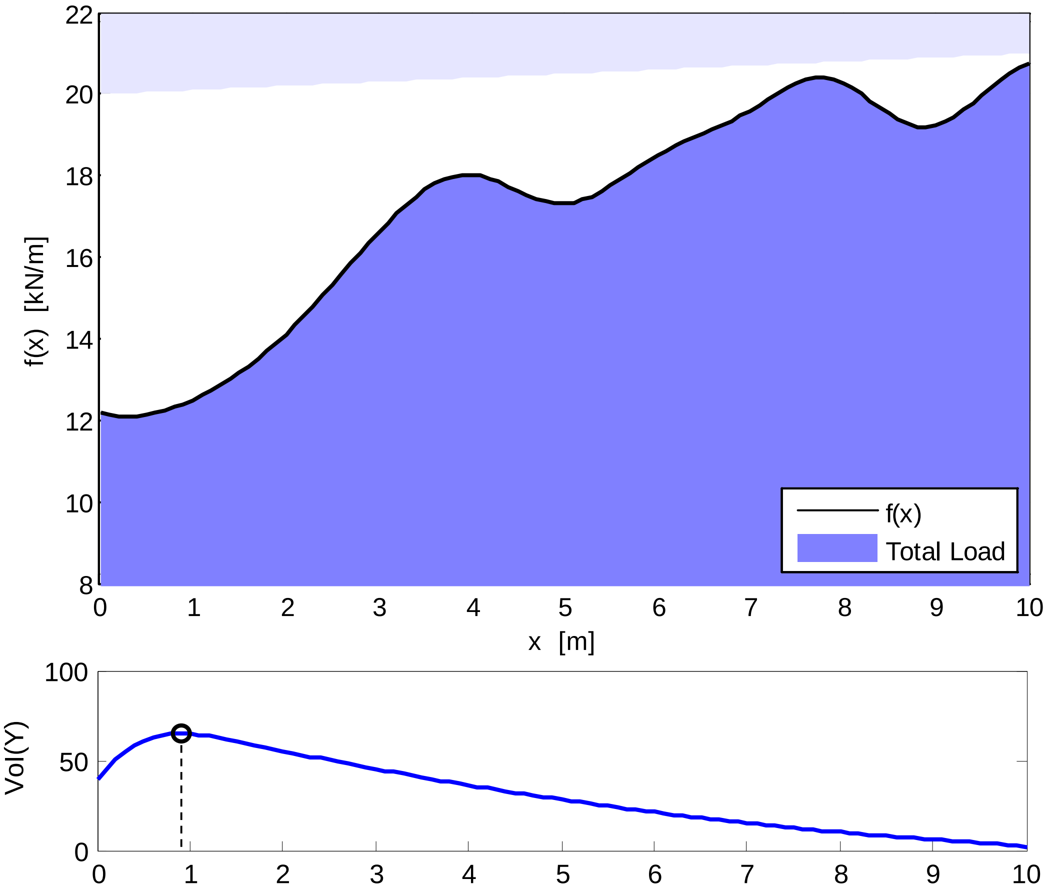
- Uncertainty in total load
- Sum of local loads

VoI for Uncertainty Reduction: Global Load

$$F = \sum_{i=1,\ldots,n} f(x_i) = \mathbf{1}^{\mathrm{T}} \boldsymbol{f}$$

$$\boldsymbol{f} = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} \sim \mathcal{N}(\boldsymbol{\mu}_f, \boldsymbol{\Sigma}_f)$$

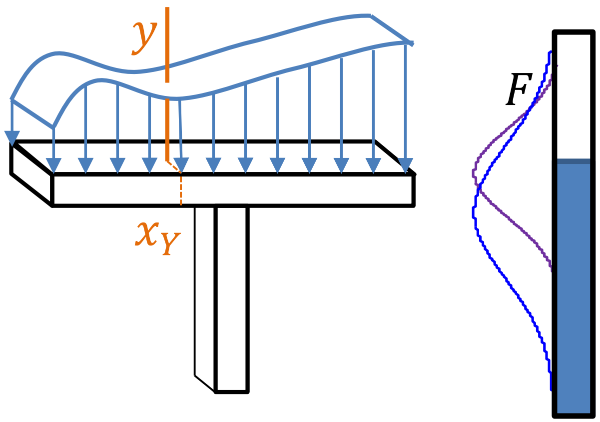
$$F \sim \mathcal{N}(\mathbf{1}^{\mathrm{T}} \boldsymbol{\mu}_f, \mathbf{1}^{\mathrm{T}} \boldsymbol{\Sigma}_f \mathbf{1})$$



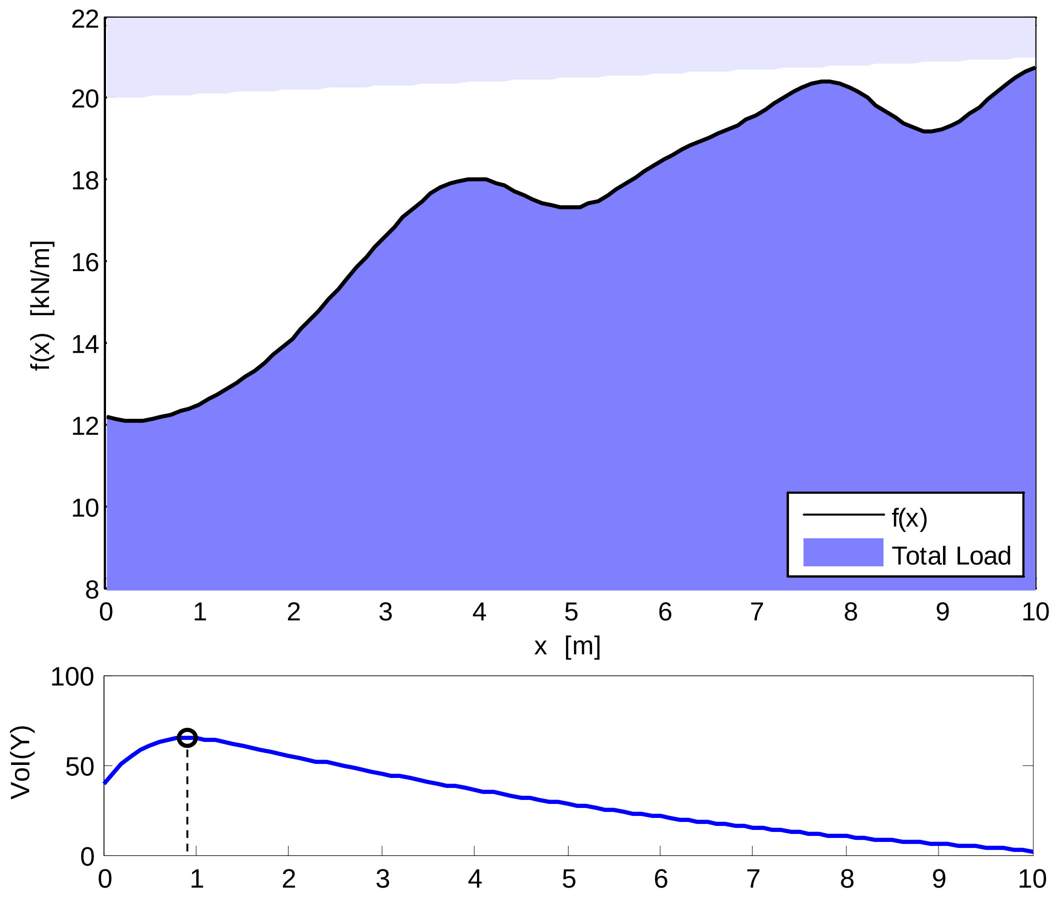
Global loading

- Uncertainty in total load
- Sum of local loads
- Univariate Gaussian

VoI for Uncertainty Reduction: Global Load



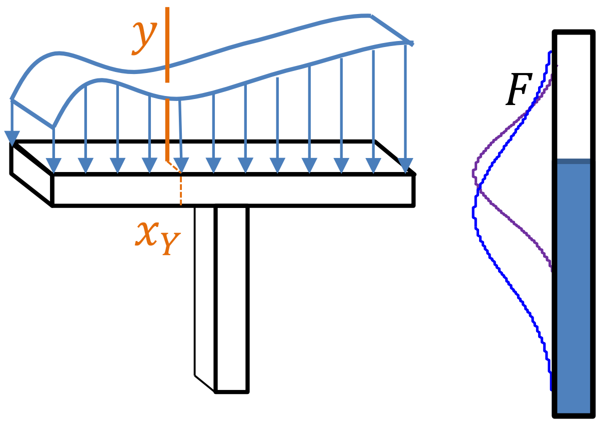
$$\mathrm{Var}[F|Y] = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\mathbf{1}$$



Global loading

- Uncertainty in total load
- Sum of local loads
- Univariate Gaussian
- Related to local uncertainty

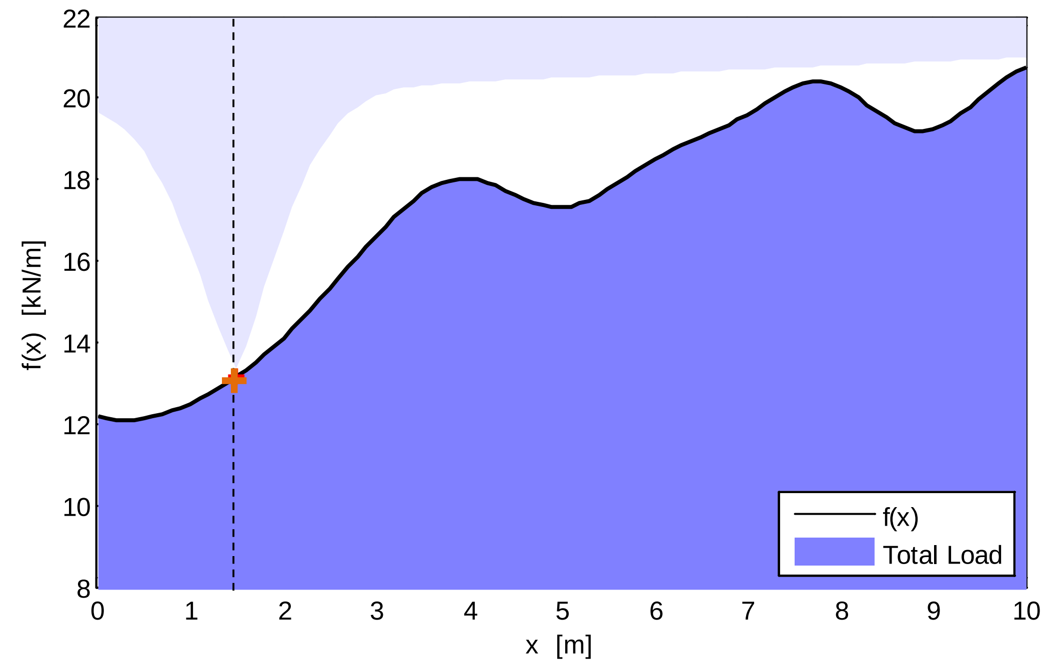
VoI for Uncertainty Reduction: Global Load

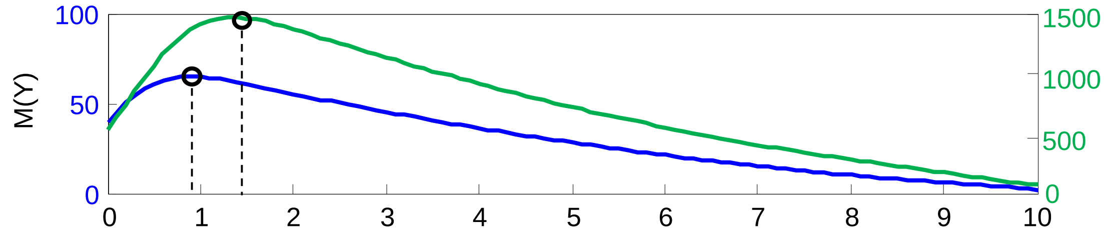


$$\mathrm{Var}[F|Y] = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\mathbf{1}$$

Global loading

- Uncertainty in total load
- Sum of local loads
- Univariate Gaussian
- Related to local uncertainty

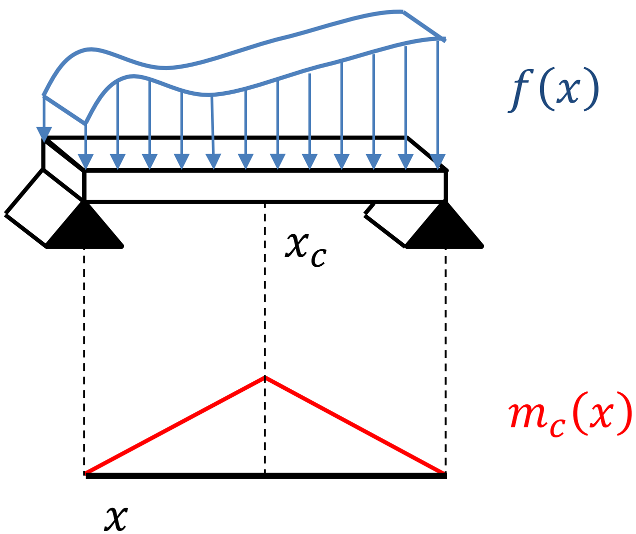


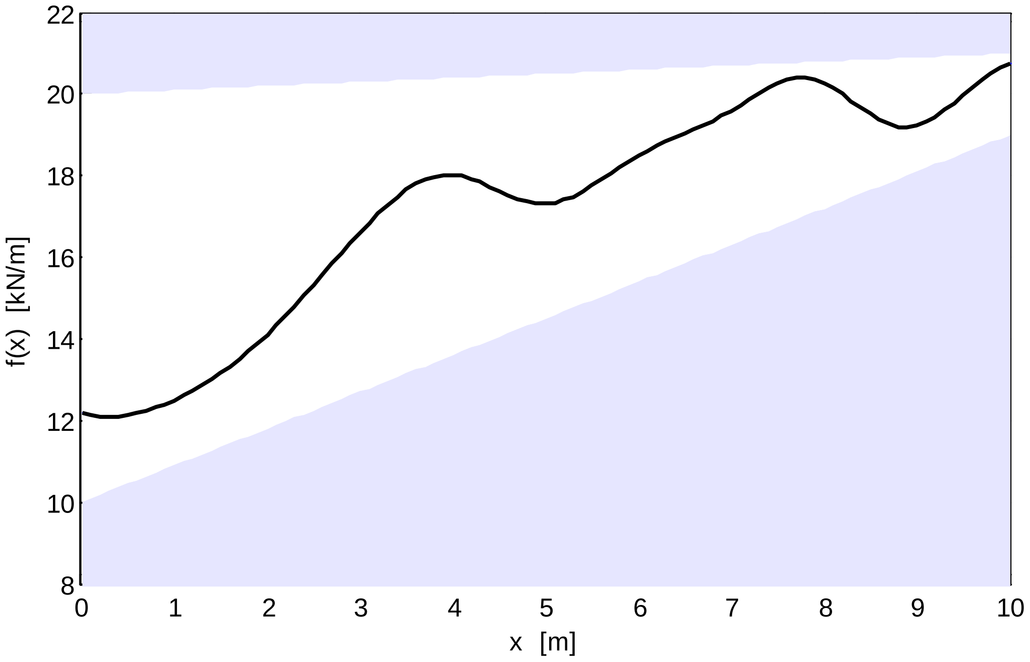


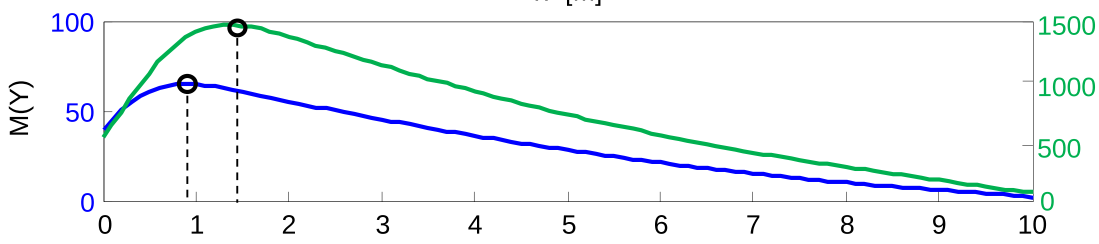
$$\mathrm{VoI}_{ERR}(Y) = \mathrm{tr}\left(\boldsymbol{\Sigma}_f\right) - \mathrm{tr}\left(\boldsymbol{\Sigma}_{f|Y}\right)$$

$$\mathrm{VoI}_{GL}(Y) = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_f\mathbf{1} - \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\mathbf{1}$$

VoI for Uncertainty Reduction: Critical Moment







Influence Lines

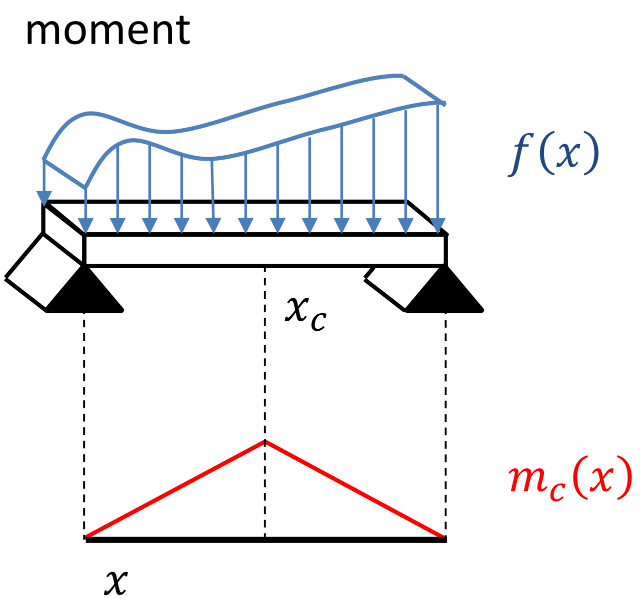
- Describe effects of local loads on reaction, shear, moment, etc.
- ex. moment at center of simply supported beam.

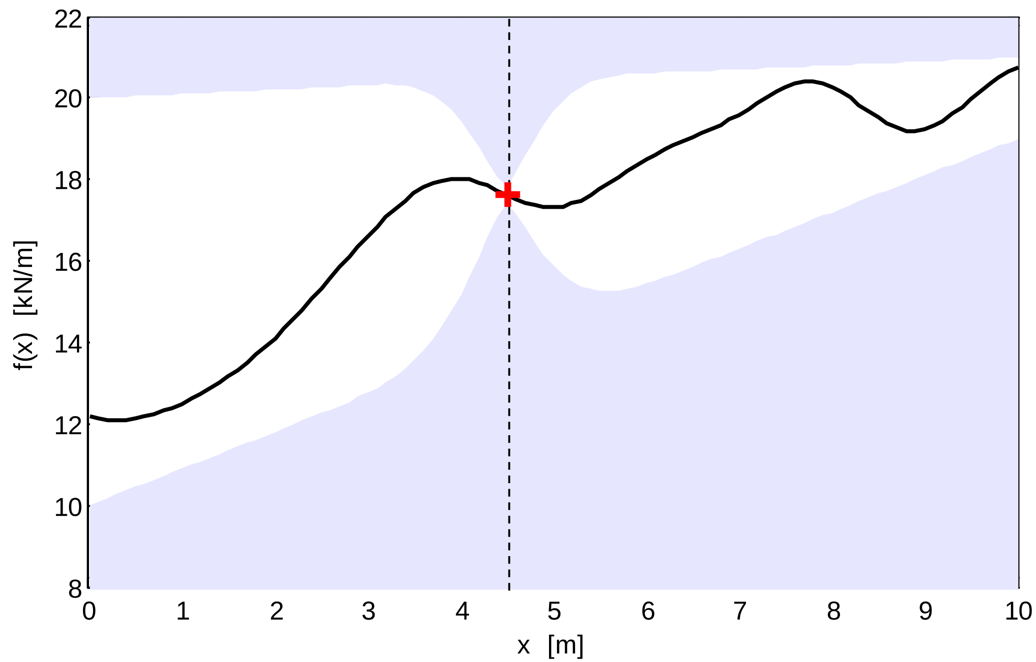
$$Moment_c = \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{f}$$

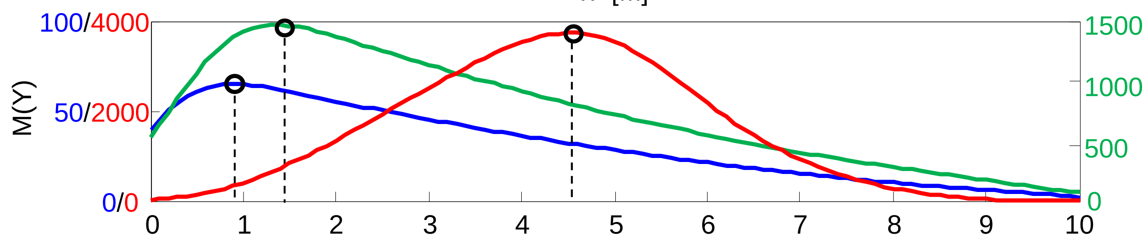
$$\mathrm{Var}[Moment_c] = \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{\Sigma}_f\boldsymbol{m}_c$$

$$\mathrm{VoI}_{Moment}(Y) = \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{\Sigma}_f\boldsymbol{m}_c - \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\boldsymbol{m}_c$$

VoI for Uncertainty Reduction: Critical Moment







Influence Lines

- Describe effects of local loads on reaction, shear, moment, etc.
- ex. moment at center of simply supported beam.

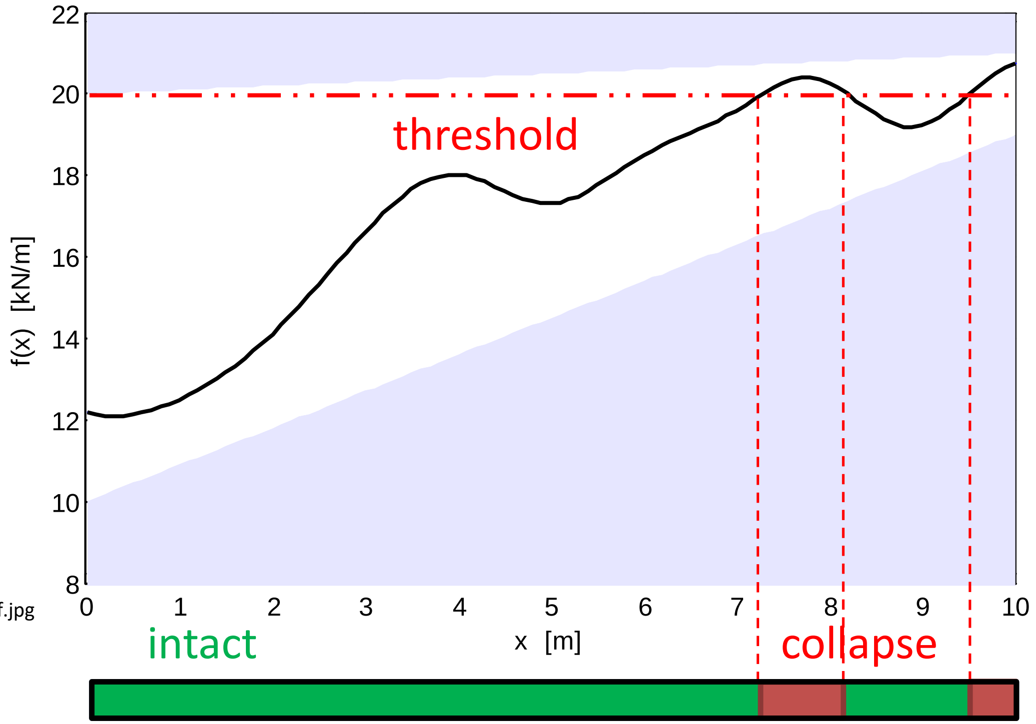
$$M_{ERR}(Y) = \mathrm{tr}\left(\boldsymbol{\Sigma}_f\right) - \mathrm{tr}\left(\boldsymbol{\Sigma}_{f|Y}\right)$$

$$M_{GL}(Y) = \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_f\mathbf{1} - \mathbf{1}^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\mathbf{1}$$

$$M_{Moment}(Y) = \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{\Sigma}_f\boldsymbol{m}_c - \boldsymbol{m}_c^{\mathrm{T}}\boldsymbol{\Sigma}_{f|Y}\boldsymbol{m}_c$$

VoI for Uncertainty Reduction: Local Failure

http://www.structuretech1.com/wp-content/uploads/2014/02/Collapsed-Kmart-Roof.jpg

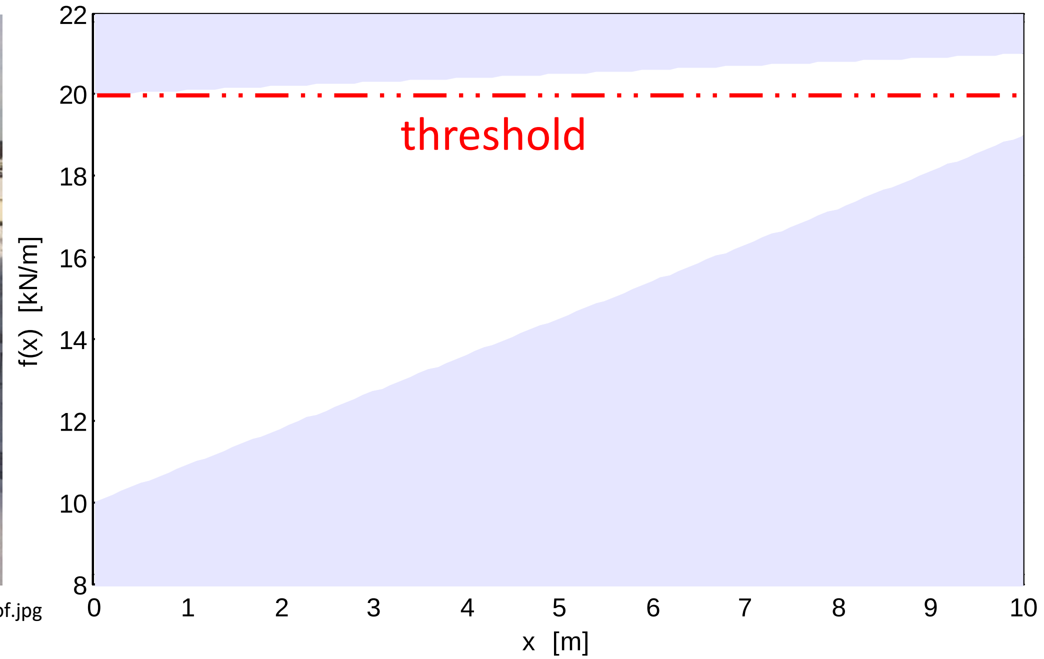


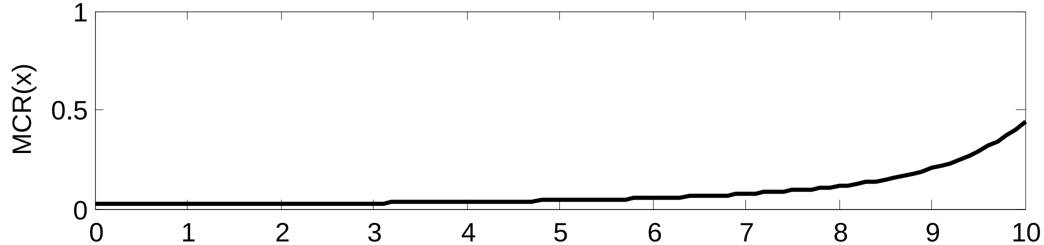
Threshold Classification

- Threshold levels distinguish discrete system states
- Example: loading exceeding roof capacity causes collapse
- Reducing uncertainty in some areas is more “important” than others

VoI for Uncertainty Reduction: Local Failure

http://www.structuretech1.com/wp-content/uploads/2014/02/Collapsed-Kmart-Roof.jpg





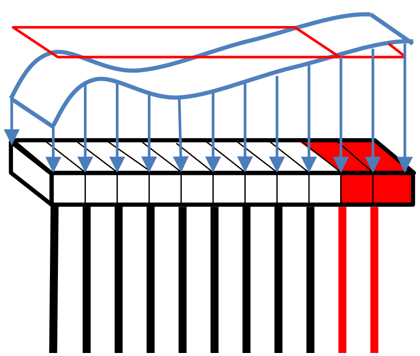
Threshold Classification

- Loss : Misclassification Rate
- State: above/below threshold
- Action: prediction of state

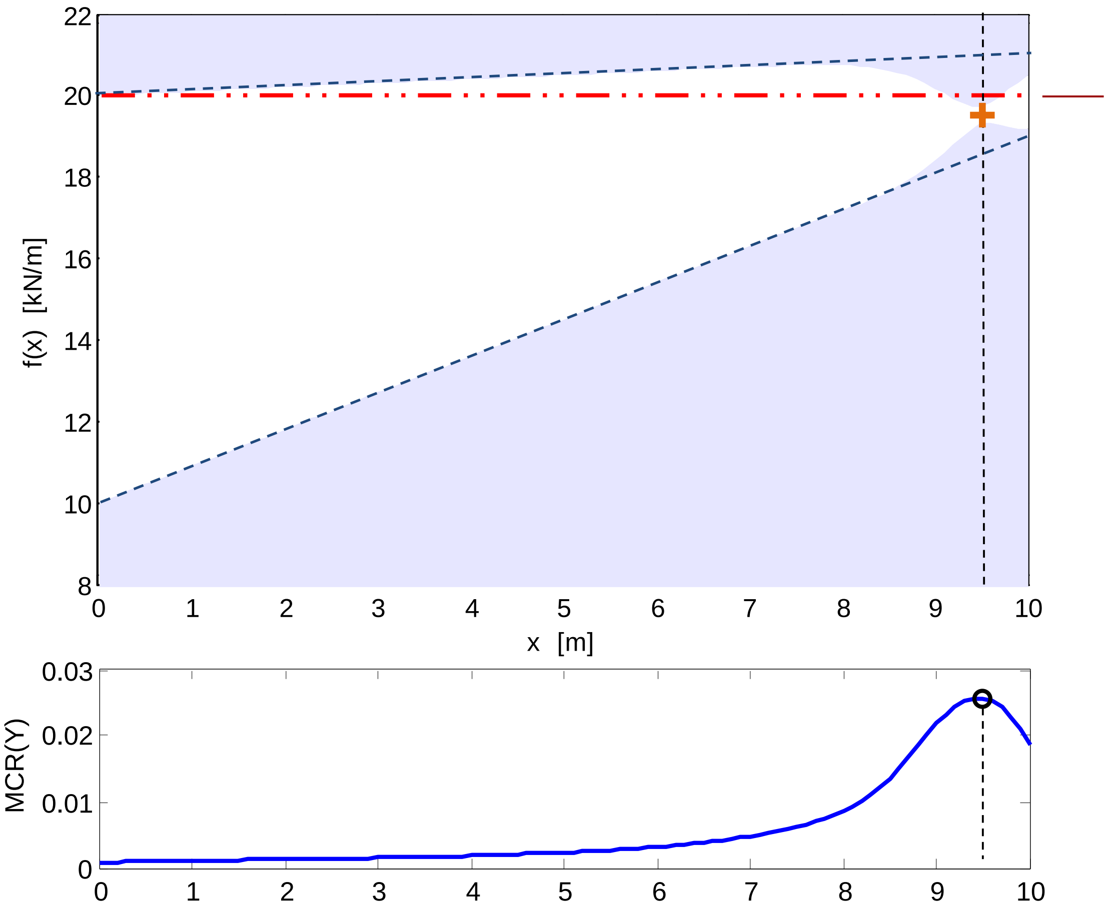
$$l(f(x), a(x)) = \begin{cases} 0 & \text{if } f(x) \leq T \text{ and } a(x) = 0 \\ 0 & \text{if } f(x) > T \text{ and } a(x) = 1 \\ 1 & \text{otherwise} \end{cases}$$

VoI for Uncertainty Reduction: Local Failure

localized failures



$$\mathrm{VoI}_{MCR}(Y) = \frac{1}{n}\sum_{i=1,\ldots,n} \mathrm{MCR}(x_i) - \mathrm{MCR}(x_i|Y)$$

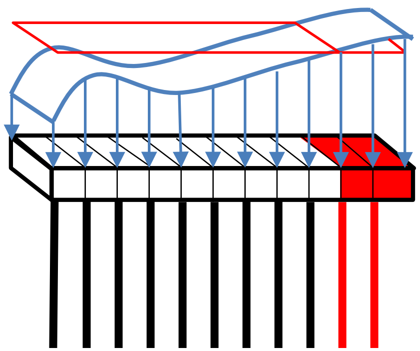


Threshold Classification

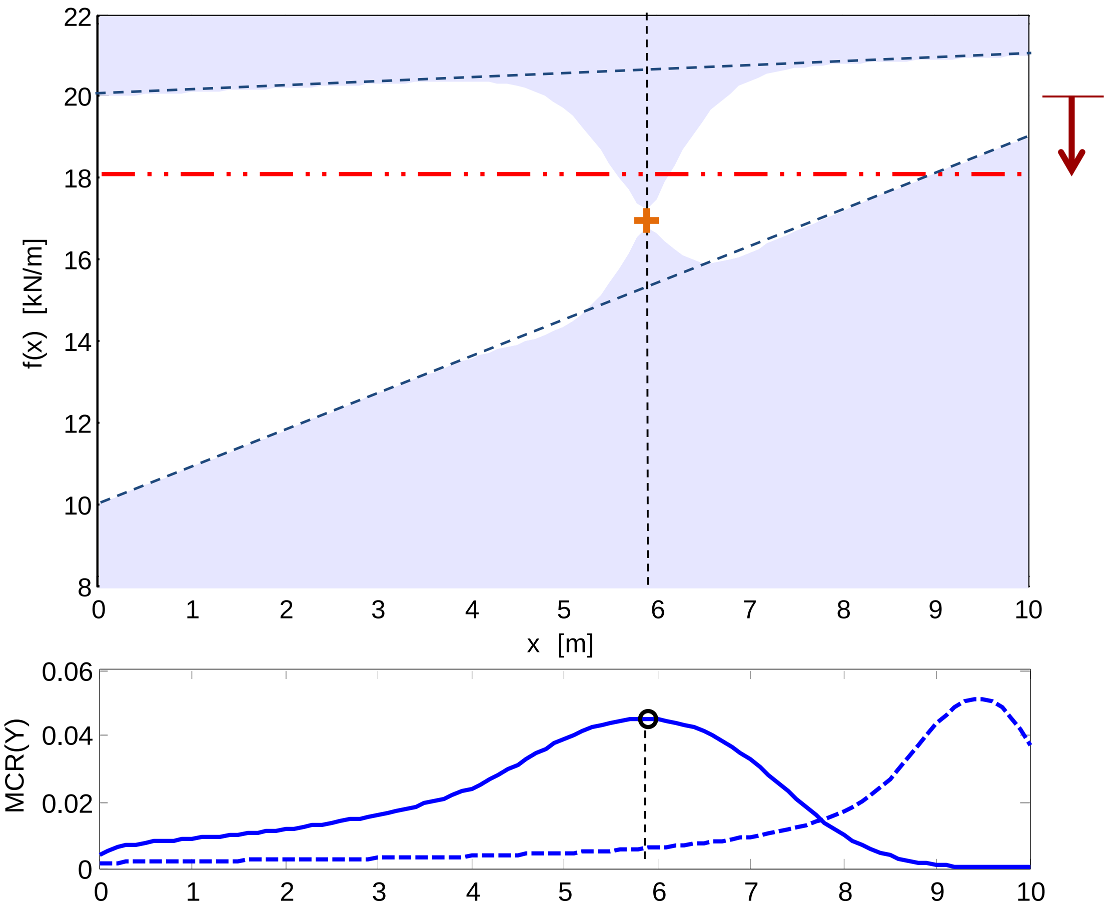
- Monitor risky areas
- Focus on uncertainty reduction for state, rather than for random field values
- Improve posterior state prediction

VoI for Uncertainty Reduction: Local Failure

localized failures



$$\mathrm{VoI}_{MCR}(Y) = \frac{1}{n}\sum_{i=1,\dots,n} \mathrm{MCR}(x_i) - \mathrm{MCR}(x_i|Y)$$



Threshold Classification

- Monitor risky areas
- Focus on uncertainty reduction for state, rather than for random field values
- Improve posterior state prediction