Value of information analysis for efficient sampling

Jo Eidsvik
Mathematical Sciences
NTNU

Pirate example

$a \in \{0,1\}$

- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

$x \in \{0,1\}$

Pirate makes decision based on preferences and maximum **utility** or **value**!

- Digging cost.
- Revenues if he finds the treasure .

$$\max_{a\in\{0,1\}} \{E(v(x,a))\}$$

Mathematics of decision situation:

- **Alternatives**

$$a \in \{0,1\} = A$$

- **Uncertainties (probability distribution)**

$$x \in \{0,1\} = \Omega \qquad p(x=1) = 0.01$$

- **Values**

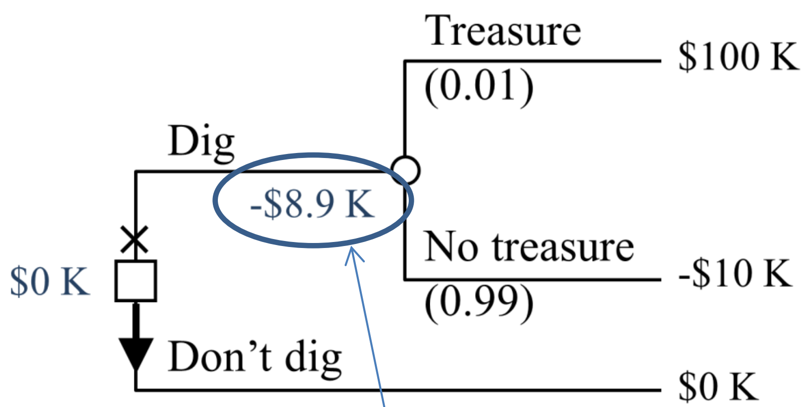
$$v = v(x,a)$$

$$v(x=0, a=1) = -10000 \qquad v(x=1, a=1) = 100000 \qquad v(x, a=0) = 0$$

- **Maximize expected value**

$$a^* = \arg\max_{a \in A} \{E(v(x,a))\}$$

Pirate's decision situation



$$E\left(u\left(v_{dig}\right)\right) = E\left(v_{dig}\right) = 0.01(100000) + 0.99(-10000) = -8900$$

Risk neutral!

Pirate example

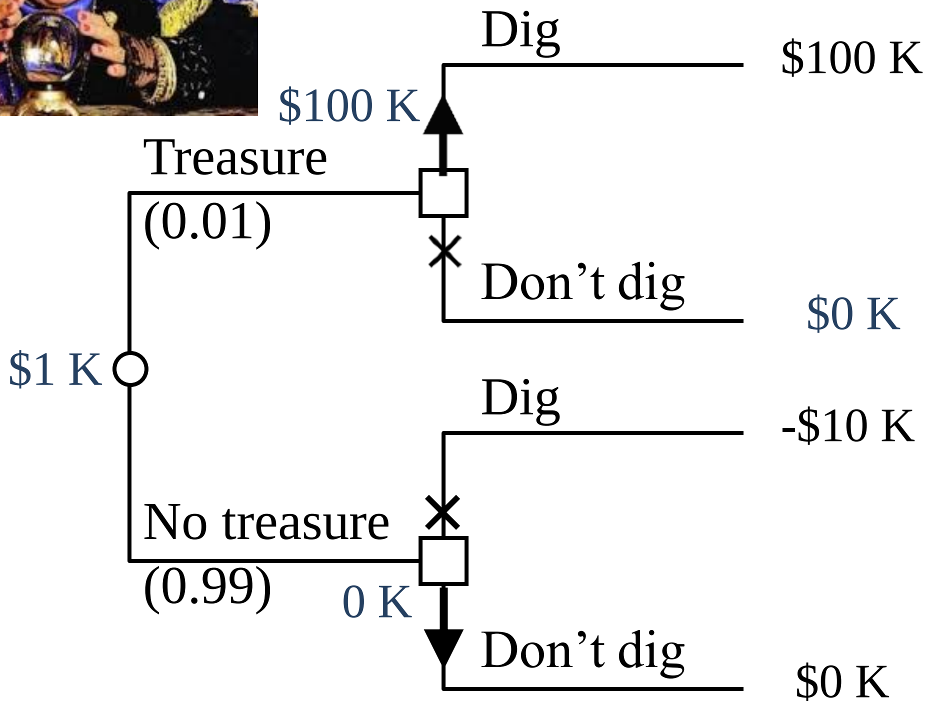
- **Pirate example**: A pirate must decide whether to dig for a treasure, or not. The treasure is absent or present (uncertainty).

- Pirate can collect **data** before making the decision, if the experiment is worth its price!

- Perfect information. Clairvoyant!

- Imperfect information. Detector!

PoV – decision tree, perfect information



$$VoI(x) = PoV(x) - PV = 1 - 0 = \$1K$$

Conclusion: Consult clairvoyant if (s)he charges less than $1000.

Which data are valuable?

Five Vs of big data:
- Volume
- Variety
- Velocity
- Veracity
- **Value**

We must acquire and process data that has value!
There is often a clear decision to make /question to answer, and data should help us.

Awareness of goal will help justify (oceanographic) observations!

VOI and spatial applications

- **Alternatives are spatial**, selection of fish sites, bridge location, release location, intervention, etc. $\boldsymbol{a}$

- **Uncertainties are spatial**; trends, dependence, multivariable processes. $\boldsymbol{x}$

- **Value function is spatial**, profit of environmental sustainability, low variance entropy, or classification error, etc.

$$v(\boldsymbol{x}, \boldsymbol{a})$$

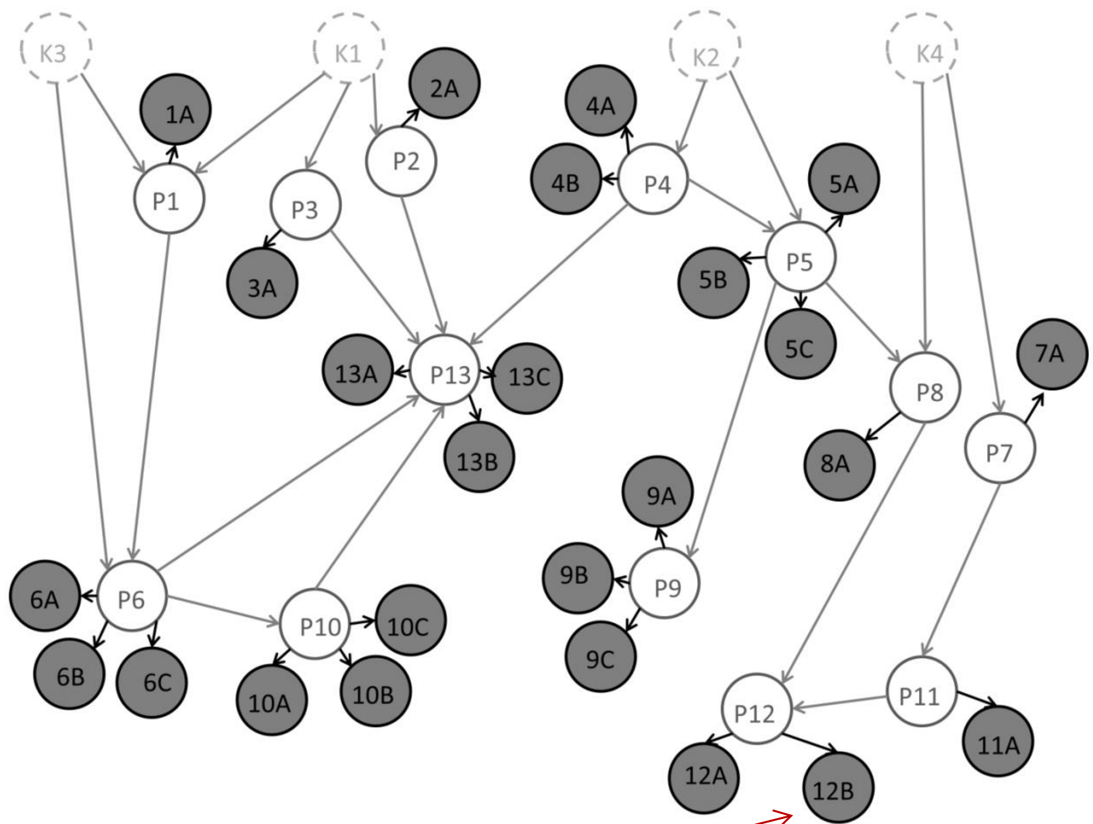
- **Data are spatial**. Surveying, processing, perfect, imperfect data, etc. $\boldsymbol{y}$

  - **What, how and where** to sample data?

  - What data are available – is that assimilated in the model ?

Uncertainty - Dependence

Gray nodes are petroleum reservoir segments where the company aims to develop profitable amounts of oil and gas.

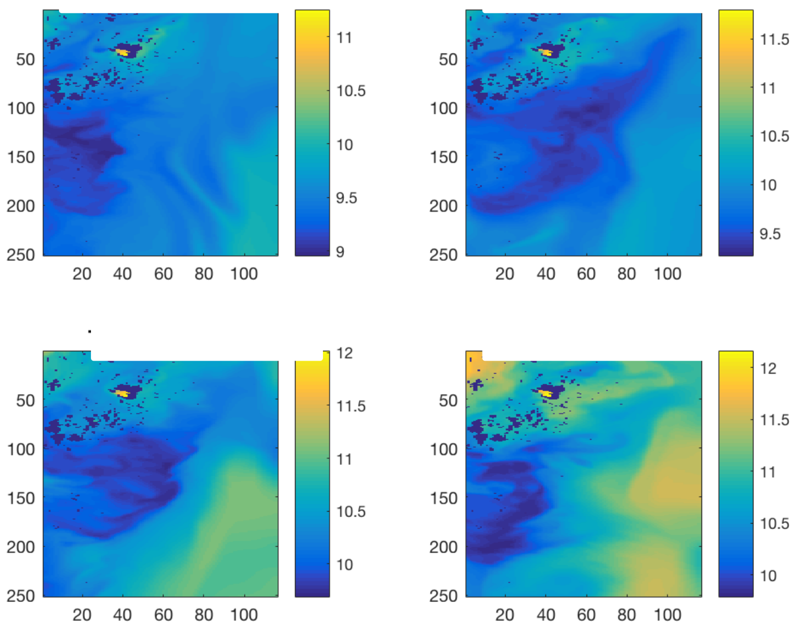


Drill the exploration well at this segment! The value of information is largest.

- **Design sampling schemes that aid decision making!**
- **Learning key uncertainties of decision situation will help under-sampling!**

Ocean temperatures – Froan area

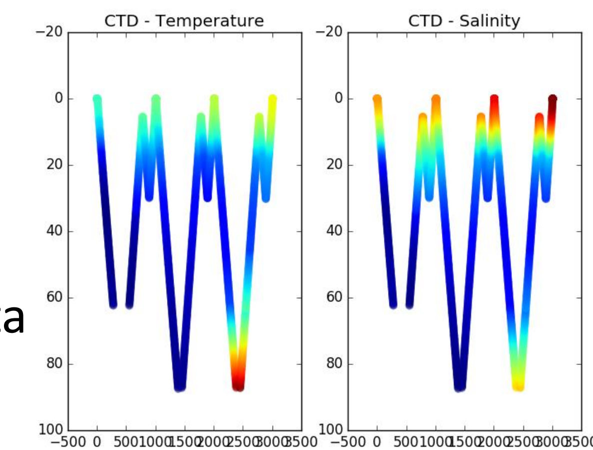
Prior realizations
(from SinMod ocean models)



Questions:

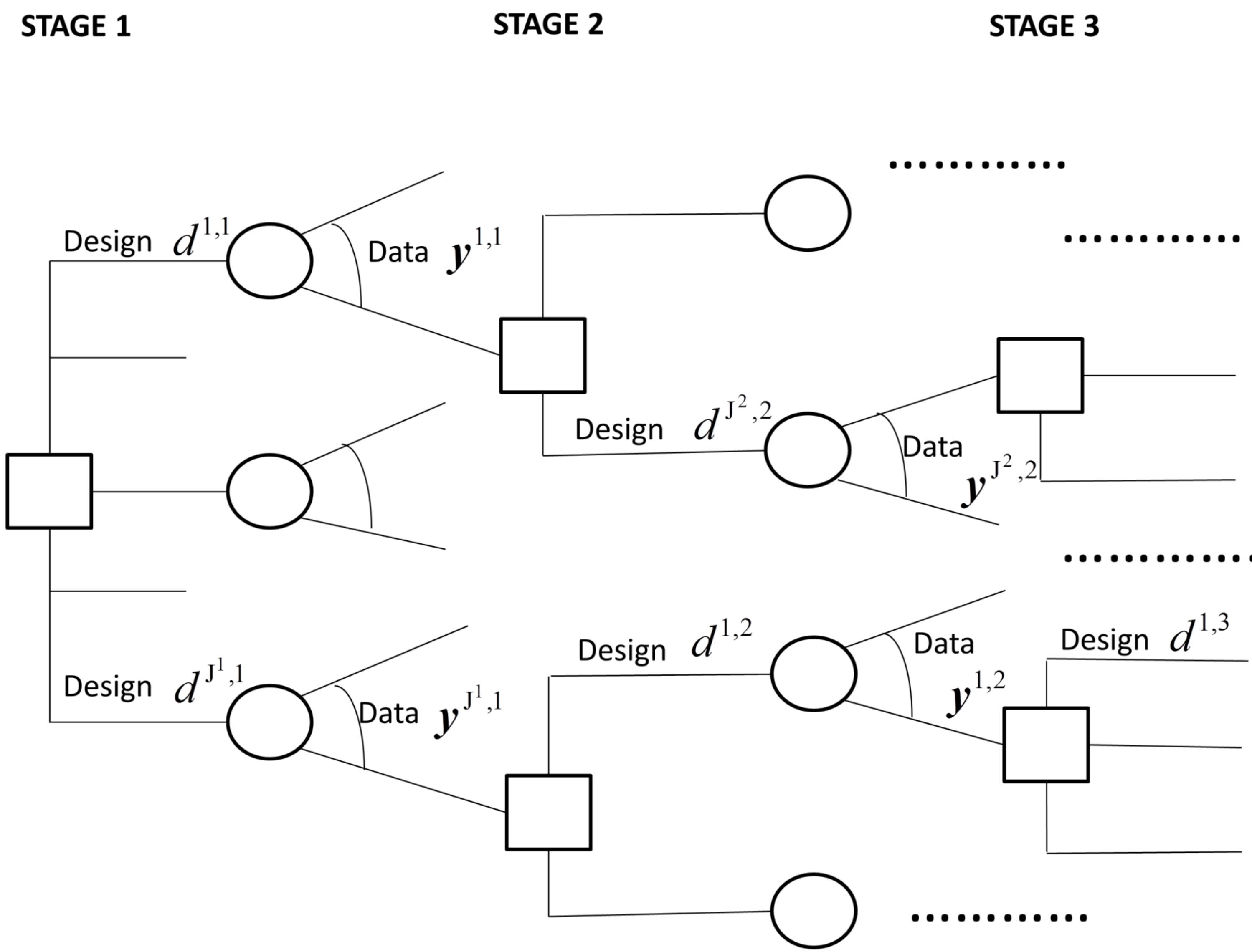
- Environmental challenges
- Fish farming
- Algae bloom
- Characterizing fronts

Typical AUV data



Select valuable AUV survey paths + Assimilate AUV data sequentially

Sequential survey path selection

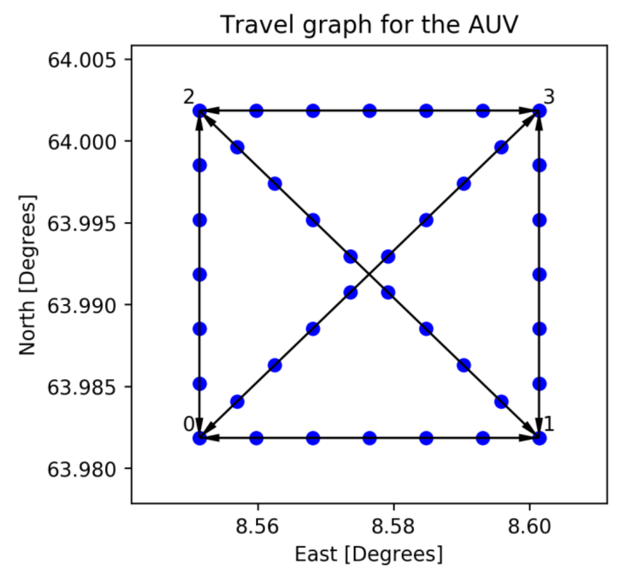


Adaptive sequential algorithm

1. Find next best survey line (if any) from information criterion.
2. Collect temperature data at selected survey line.
3. Update temperature model (Gaussian process) in spatial domain with survey data.
4. Go to 1.

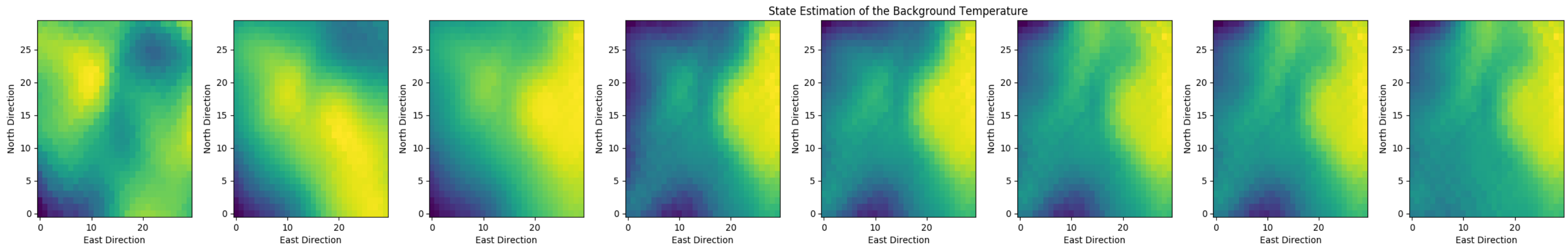
Greedy/myopic heuristic for dynamic program.

Goal:
- Find high temperature gradients
- Reduce prediction variance in temperature

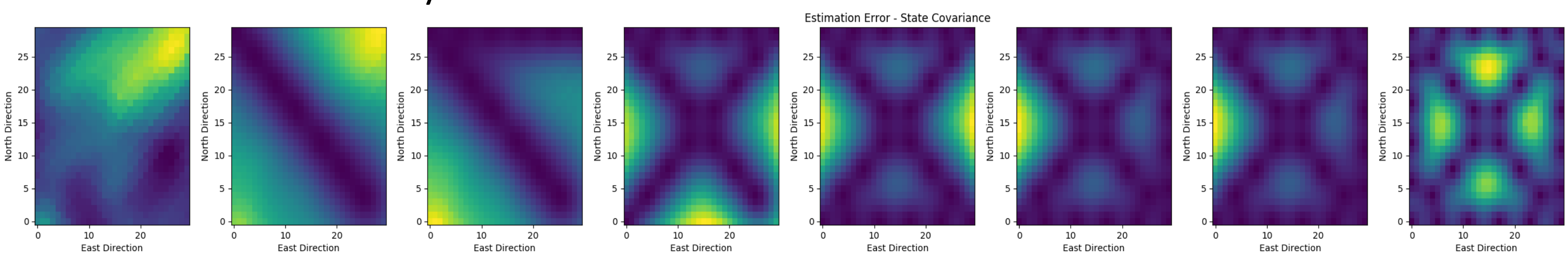


Results of adaptive algorithm

Mean of one survey realization



Variance of one survey realization



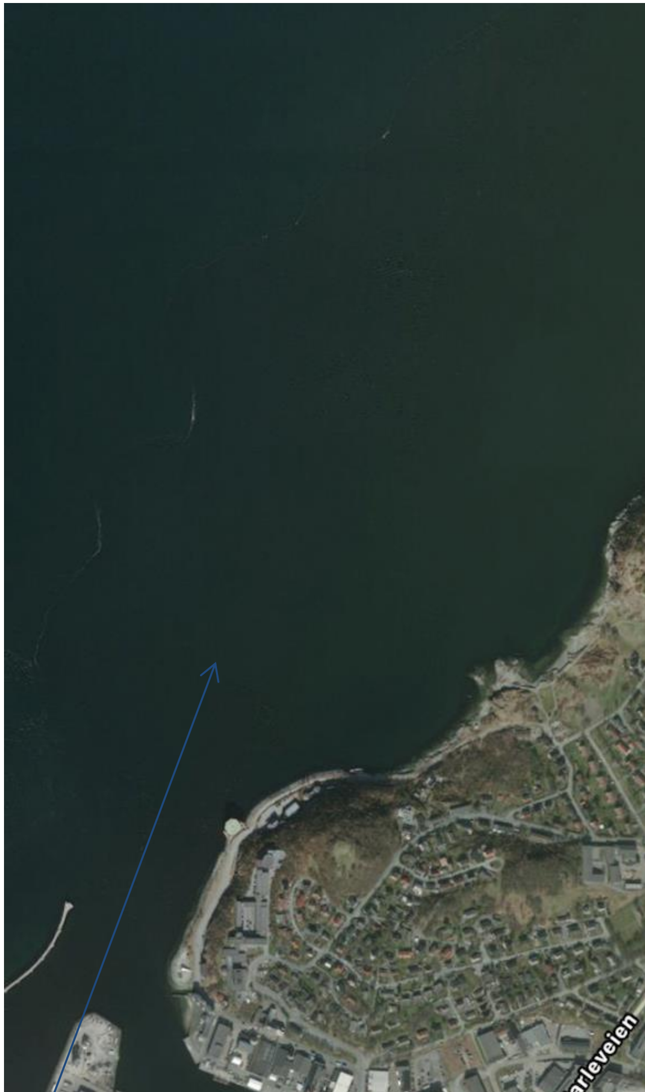
$$\boldsymbol{x} \sim GP\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$$

$$\boldsymbol{y} = \boldsymbol{A}_d \boldsymbol{x} + N\left(\boldsymbol{0}, \boldsymbol{T}_d\right)$$

$$\boldsymbol{\mu} = \boldsymbol{\mu} - \boldsymbol{\Sigma} \boldsymbol{A}_d^t \left(\boldsymbol{A}_d \boldsymbol{\Sigma} \boldsymbol{A}_d + \boldsymbol{T}_d\right)^{-1} \left(\boldsymbol{y} - \boldsymbol{A}_d \boldsymbol{\mu}\right)$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma} - \boldsymbol{\Sigma} \boldsymbol{A}_d^t \left(\boldsymbol{A}_d \boldsymbol{\Sigma} \boldsymbol{A}_d + \boldsymbol{T}_d\right)^{-1} \boldsymbol{A}_d \boldsymbol{\Sigma}$$

Fresh cold water & salt warm water



River
mouth

River
mouth

Excursion sets and excursion probabilities

$$ES_a = \{\mathbf{s} : x(\mathbf{s}) < a\}$$

Connections to active learning.

$$\boldsymbol{x} \sim GP(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$EP_a(\mathbf{s}) = P(x(\mathbf{s}) < a)$$

$$\boldsymbol{y}_d = \boldsymbol{A}_d \boldsymbol{x} + N(\mathbf{0}, \boldsymbol{T}_d)$$

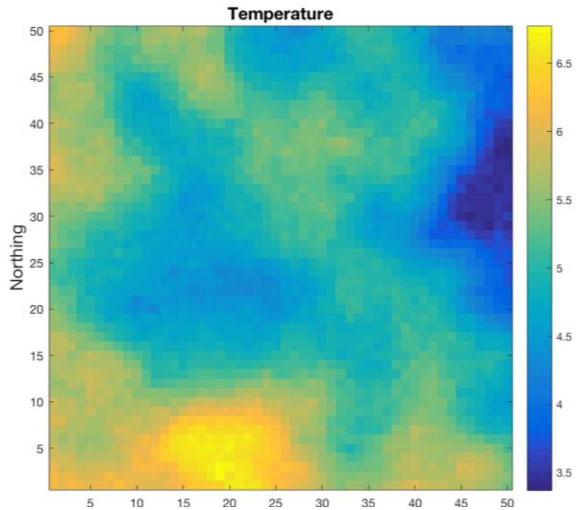
$$EP_a(\mathbf{s} \mid \boldsymbol{y}_d) = P(x(\mathbf{s}) < a \mid \boldsymbol{y}_d)$$

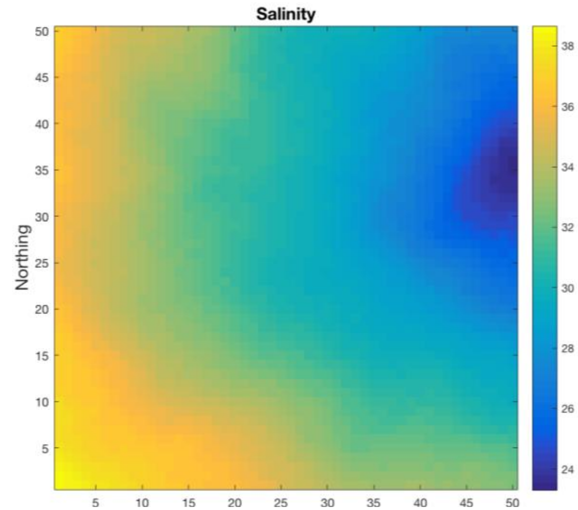
Criterion for path selection:

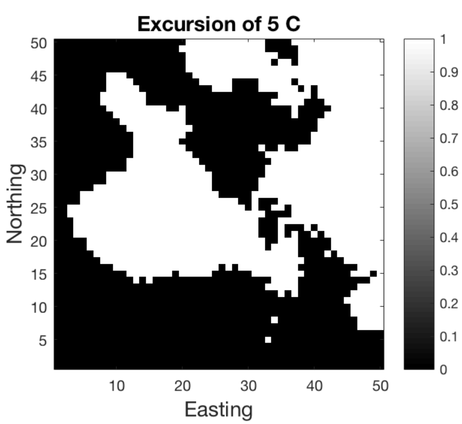
$$d^* = \arg\min_d \left\{ \int\int \underbrace{EP_a(\mathbf{s} \mid \boldsymbol{y}_d)\, p(\boldsymbol{y}_d)\, d\boldsymbol{y}_d}_{} \, d\mathbf{s} \right\}$$

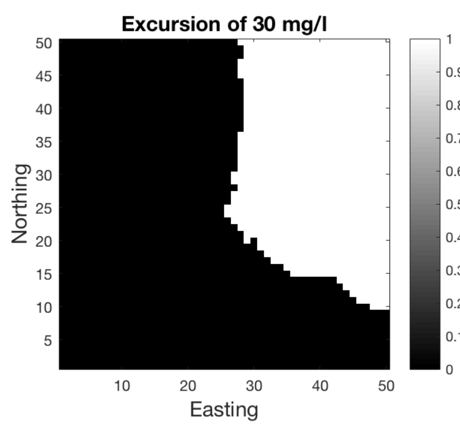
Closed form for Gaussian processes.

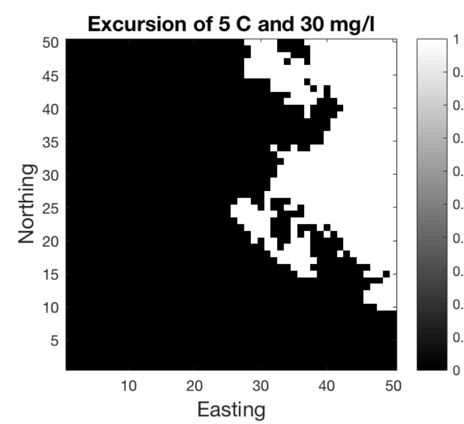
Bivariate excursion sets



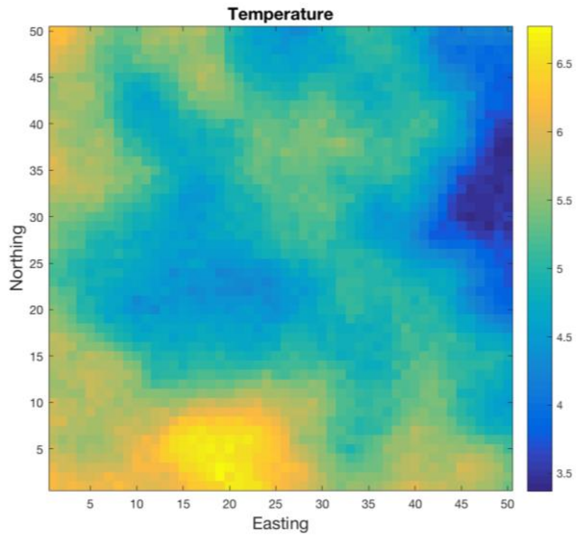


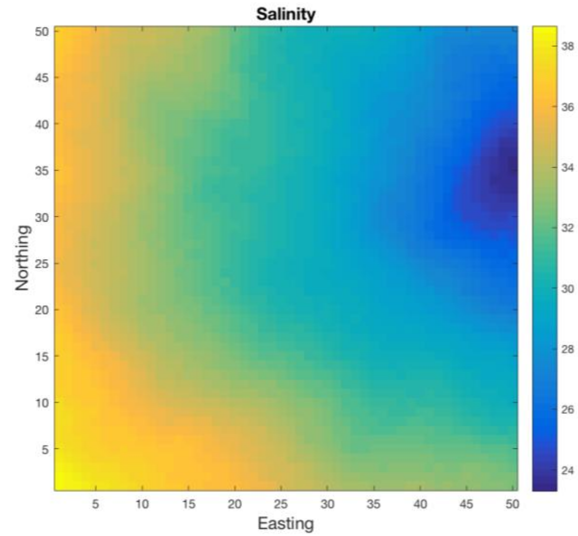






Bivariate excursion sets – closed form





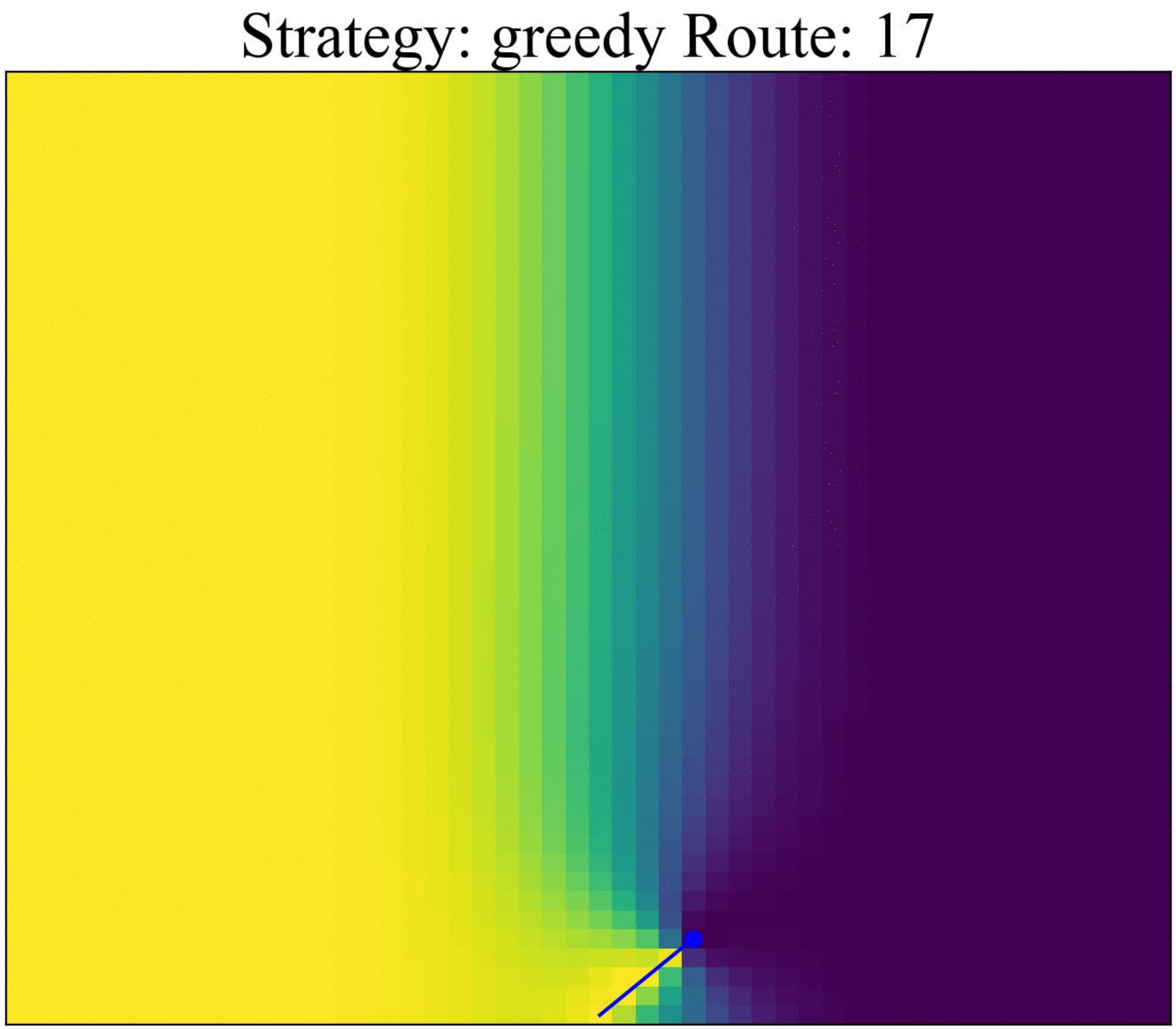
$$\int EP_a(\boldsymbol{s} \mid \boldsymbol{y}_d)\, p(\boldsymbol{y}_d)\, d\boldsymbol{y}_d = \int P\left(x_{\text{temp}}(\boldsymbol{s}) < \alpha_{\text{temp}}, x_{\text{sal}}(\boldsymbol{s}) < \alpha_{\text{sal}} \mid \boldsymbol{y}_d\right) p(\boldsymbol{y}_d)\, d\boldsymbol{y}_d$$

$$= \Phi_2(\boldsymbol{0}_2, \boldsymbol{\beta}_2, \boldsymbol{R}_2) - \Phi_4(\boldsymbol{0}_4, \boldsymbol{\beta}_4, \boldsymbol{R}_4)$$

Multivariate Gaussian cumulative distribution function

Standard matrix –vector computations.

Myopic path selection for excursions



Real time excursion probability (blue = cold fresh water, yellow = salt warm water.

References

Chevalier et al. Fast parallel kriging-based stepwise uncertainty reduction with appication to the identification of an excursion set, Technometrics, 2014.

Cressie, N. and Wikle, C., Statistics for spatio-temporal data, Wiley, 2011.

Eidsvik, J., Mukerji, T. and Bhattacharjya, D., Value of information in the Earth sciences, Cambridge University Press, 2015.

Fossum, T.O. et al., Information-driven robotic sampling in the coastal domain, Journal of field robotics, 2018.

Howard R.A. and Abbas, A.E., Foundations of decision analysis, Pearson, 2015.