# Gaussian Process for Excursion Set Estimation: Example Applications

Cedric Travelletti

*cedric.travelletti@stat.unibe.ch*

January 10, 2022

Joint work with D. Ginsbourger, N. Linde, J. Eidsvik, K. Rajan and T. O. Fossum

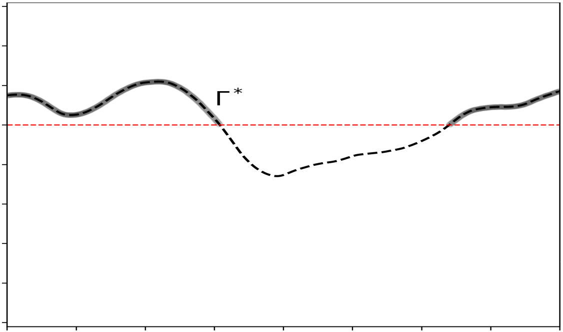
# Excursion Set Estimation

## Goal

Estimate excursion set

$$\Gamma^* := \{x \in D : \; f(x) > T\} \, ,$$

where $f : D \to \mathbb{R}$ is some unknown target function.

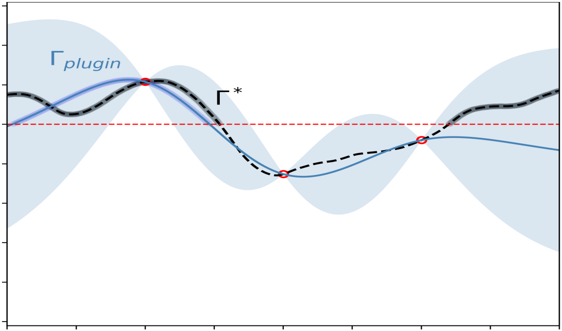



(a) Target function $f$ and true excursion set.

# Bayesian Excursion Set Estimation

**Bayesian approach:** $f$ is a realization of a GP $(Z_x)_{x \in D}$.

- Use posterior to approximate true $\Gamma^*$.




(a) Target function $f$, GP posterior, excursion set and plugin estimate.

# Extending to Inverse Problems

GPs can also handle linear operator data:
$y_i = G_i(f) = \int_D f(x) d\lambda_i(x)$, where $G_i : C(D) \to \mathbb{R}$ is a linear form.

$$
\begin{aligned}
m_n(x) &= m(x) + K_{x\boldsymbol{G}_n} K_{\boldsymbol{G}_n \boldsymbol{G}_n}^{-1} \left(\boldsymbol{y}_n - \boldsymbol{G}_n m.\right) \\
k_n(x, x') &= k(x, x') - K_{x\boldsymbol{G}_n} K_{\boldsymbol{G}_n \boldsymbol{G}_n}^{-1} K_{x'\boldsymbol{G}_n}^T \\
K_{x\boldsymbol{G}_n} &= \left( \int_D k(x, y)\, d\lambda_i(y) \right)_{i=1,\dots,n} \\
K_{\boldsymbol{G}_n \boldsymbol{G}_n} &= \left( \int_D \int_D k(y, z)\, d\lambda_i(y)\, d\lambda_j(z) \right)_{i,j=1,\dots,n}
\end{aligned}
$$

Travelletti, C., Ginsbourger, D., Linde, N. (2021). *Uncertainty quantification and experimental design for large-scale linear inverse problems under gaussian process priors.* arXiv preprint arXiv:2109.03457.

# Example 1: Estimating High-Density Regions inside a Volcano

# Example: Stromboli Volcano

Want to learn the interior structure of the Stromboli volcano.

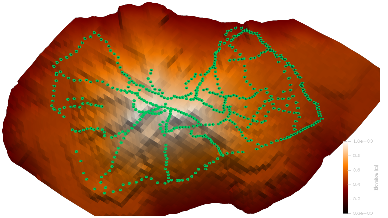
- Only allowed to measure gravitational field on the surface

**Question**

Where should I collect data to get the best possible reconstruction?

# Problem Setup

- **Want to recover:** unknown density field $\rho : D \to \mathbb{R}$
- measurement sites $s_1, \dots, s_n \in S$ on the surface
- **Available data:** surface gravity field $\{G_{s_i}[\rho]\}_{i=1,\dots,n}$.

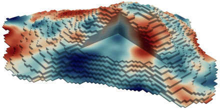


## Measurement (forward) operator

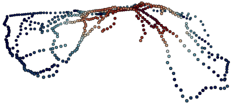
$$G_{s_i}[\rho] = \int_D \rho(x) \frac{x^{(3)} - s_i^{(3)}}{\|x - s_i\|^3} dx$$

# Excursion Set Estimation

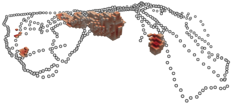
- Want to recover high-density regions.
- Correspond to geological features of interest.
- Propose *adaptive* data collection plan.



(a) density



(b) observation sites (static design)



(c) excursion set

# Stepwise Uncertainty Reduction

Want to learn high-density region: $\Gamma^* := \{x \in D : \ \rho(x) > T\}$

- Select observations by sequentially optimizing an uncertainty reduction criterion (SUR framework).

**targeted IMSE**

$$J_n^{tIMSE}(s) = \mathbb{E}_n\left[\int_D k_{n+1}(u,u)p_{n,T}(u)\mu(du)|G_{n+1} = G_s\right]$$

- Future variance $k_{n+1}(u,u)$ does not depend on observations.

# Experiments

- Train GP model on field campaign data.
- Generate *synthetic* ground truths by sampling the GP.
- Evaluate criterion *in-silico*.

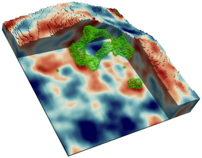


Figure: True density and excursion set (generated from model).

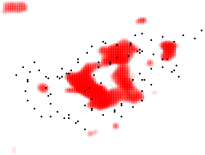


Figure: Proposed data collection plan, wIVR long-range strategy, total budget of 90 observations.

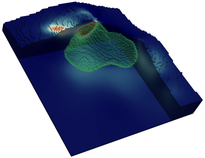


Figure: Estimated excursion set (Vorob'ev Expectation) and coverage function.

# Results

Get data collection plan by myopic optimization of the criterion.

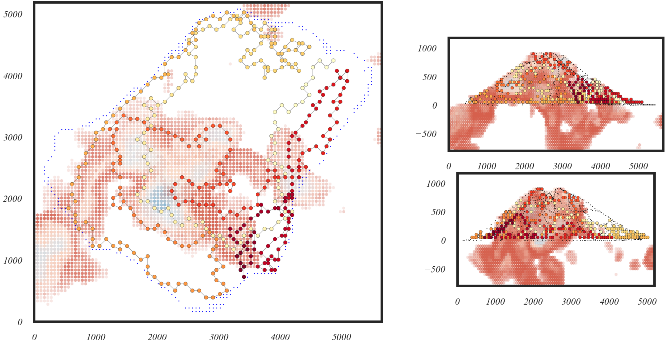
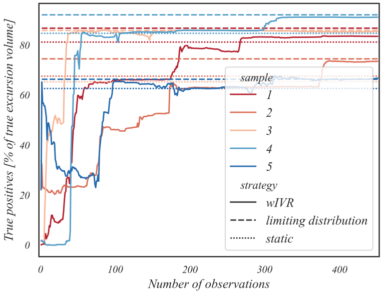


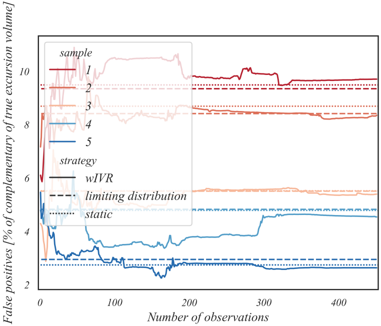

Figure: Excursion set and proposed data collection plan (tIMSE).

# Results (contd.)

- Compare proposed design to static one and to space-filling design.
- Able to reach close-to-minimal uncertainty after 250 observations.




(a) True positives




(b) False positives

Figure: Evolution of true and false positives as a function of the number of observations.

# UQ on Excursion Volume

Sample from posterior volume distribution on 5 different ground truths.

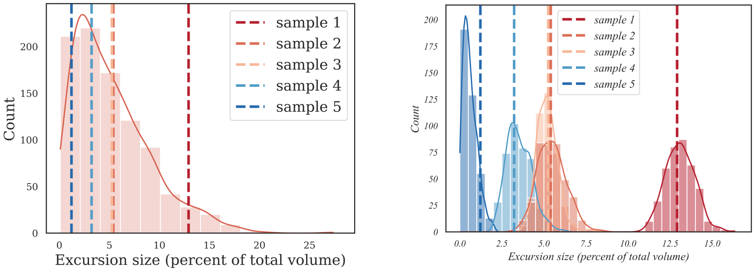


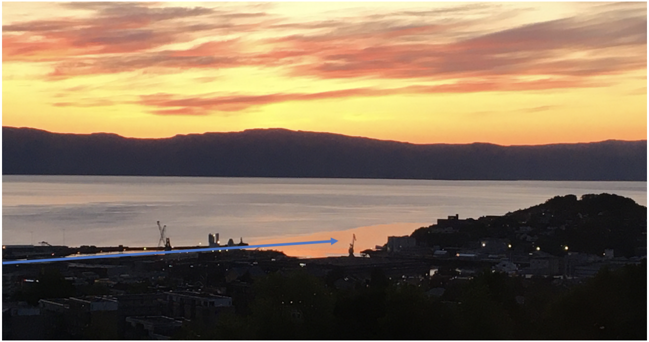

Figure: Prior (left) and empirical posterior (right) distribution (after 450 observations) of the excursion volume for each ground truth. True volumes are denoted by vertical lines.

Demonstrate uncertainty reduction and peaking around true volume.

# Example 2: River Plume Mapping

# Case Study Mapping a River Plume in a Fjord

Frontal zone where river (cold freshwater) enters ocean (warm, salty water) is home to many important biological processes.



(a) Nidelva river entering Trondheim Fjord.

Can we map the interface?

Photograph: courtesy of J. Eidsvik

# Mapping the Interface of a River Plume

River differs from ocean by its temperature and salinity content.

**Ideally: Collect temperature and salinity data only along the interface.**

But ...

- Boundary evolves (tides, wind, river discharge variations, ...).
- Need for adaptive data collection plans.

# Autonomous Data Collection

Collect temperature / salinity data using an autonomous submarine (AUV).


- AUV uses model of river plume to guide data collection process.

Use Multivariate GP to model temperature-salinity field.

# Multivariate GP Model

- River plume differs from ocean water by its low temperature and salinity.
- Model temperature and salinity as a bivariate random field.

$$Z := (Z_T, Z_S)^T \sim \mathrm{Gp}(\mu, k)$$

## Ocean is then an Excursion Set

$$\Gamma^* := \{u \in D : Z_T \geq t_T, Z_S \geq t_s\}$$

- As UAV gathers observations, GP model is updated by co-kriging.
- Kriging-equations can be made form-invariant across all dimensions of the output [FTE$^+$21].

# Multivariate GP Model (contd.)

Prior knowledge about river flow encoded in mean function of the GRF.

$$m\left(u\right) = \mathbb{E}\left[\begin{pmatrix} Z_S \\ Z_T \end{pmatrix}\right] = \begin{pmatrix} \beta_{0,S} \\ \beta_{0,T} \end{pmatrix} + \begin{pmatrix} \beta_{1,S}^{(1)} & \beta_{1,S}^{(2)} \\ \beta_{1,S}^{(1)} & \beta_{1,T}^{(2)} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

Spatial correlations and cross-correlation between tempearture and salinity modeled by a separable covariance function

$$\mathrm{Cov}\left(Z_{s,i}, Z_{u,j}\right) = k(s,u)\gamma(i,j),\ \gamma(i,j) = \begin{cases} \sigma_l^2,\ i = j \\ \gamma_0 \sigma_i \sigma_j,\ i \neq j \end{cases}$$

# Multivariate GP Model (contd.)

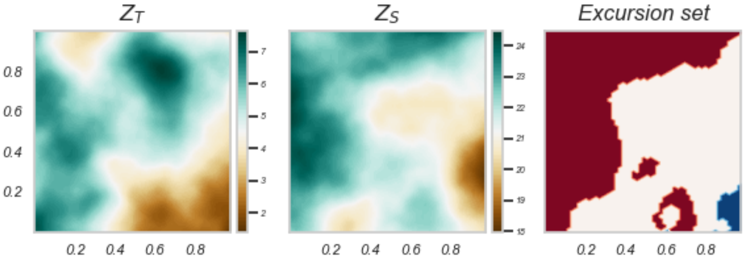
Prior knowledge about river flow encoded in mean function of the GRF.

$$m\left(u\right)=\mathbb{E}\left[\begin{pmatrix}Z_S\\ Z_T\end{pmatrix}\right]=\begin{pmatrix}\beta_{0,S}\\ \beta_{0,T}\end{pmatrix}+\begin{pmatrix}\beta_{1,S}^{(1)} & \beta_{1,S}^{(2)}\\ \beta_{1,S}^{(1)} & \beta_{1,T}^{(2)}\end{pmatrix}\begin{pmatrix}u_1\\ u_2\end{pmatrix}$$

Spatial correlations and cross-correlation between tempearture and salinity modeled by a separable covariance function

$$\mathrm{Cov}\left(Z_{s,i},Z_{u,j}\right)=k(s,u)\gamma(i,j),\ \gamma(i,j)=\begin{cases}\sigma_l^2,\ i=j\\ \gamma_0\sigma_i\sigma_j,\ i\neq j\end{cases}$$

# Example Realisation



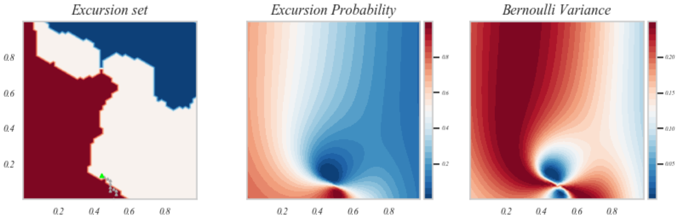

(a) Realisation of the GRF model. Excursion set in red, excursion of a single component (temperature or salinity) in white.

# Uncertainty Measures on ES

(Current) law of the field gives information about probable location of excursion set $\Gamma$.

## Coverage Function

$$p(u) = \mathbb{P}\left(Z_{T,u} \geq t_T, Z_{S,u} \geq t_s\right)$$




(a) Realisation of excursion set from the GRF model (excursion set in red). Excursion probability and Bernoulli variance conditional on data collected at locations depicted by triangles.

# Integrated Bernoulli Variance

Given excursion set $\Gamma = \{u \in D : Z_{T,u} \geq t_T, Z_{S,u} \geq t_s\}$, a useful measure of uncertainty is the Integrated Bernoulli Variance

$$H_n^{\Gamma} = \int_D p_n(u)\left(1 - p_n(u)\right) d\mu(u)$$

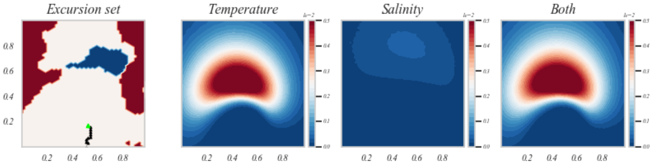
where $\mathbb{P}$ denotes the current law of the field.

- Sequential uncertainty reduction strategy by minimizing expected future uncertainty $J_n^{\Gamma}(x) = \mathbb{E}_n\left[H_{n+1}^{\Gamma} | x_{n+1} = x\right]$.
- Criterion can be computed analytically.

Fossum, T., O., Travelletti, C., Eidsvik, J., Ginsbourger, D. (2021). *Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling.* The Annals of Applied Statistics **15**, no. 2.

# Sequential Uncertainty Reduction (SUR) Strategy

At each step, choose next observations location so as to maximally decrease expected Bernoulli variance (myopic).




(a) Expected Bernoulli Variance Reduction associated to data collection at a given location (green triangle). Also shows EBV reduction associated to observing a single component of the field.

$$x_{n+1} = \underset{x \in \mathcal{J}}{\operatorname{argmin}}\, J_n^{\Gamma}(x)$$

# Simulation Study

Figure: Example run of myopic startegy. Excursion set in red. Locations where only one component of the field is above threshold are shown in pink. Radar displaying EIBV of neighbors at each step.

# References I


- Trygve Olav Fossum, Cédric Travelletti, Jo Eidsvik, David Ginsbourger, and Kanna Rajan, *Learning excursion sets of vector-valued Gaussian random fields for autonomous ocean sampling*, The Annals of Applied Statistics **15** (2021), no. 2, 597 – 618.
- Cédric Travelletti, David Ginsbourger, and Niklas Linde, *Uncertainty quantification and experimental design for large-scale linear inverse problems under gaussian process priors*, arXiv preprint arXiv:2109.03457 (2021).