Bayesian Inversion for Excursion Set Recovery: Including known Trends and Perspectives toward Sequential Uncertainty Reduction

Cedric Travelletti

October 21, 2022

Joint work with D. Ginsbourger (UniBe), N. Linde (UniL) and A. Gautier (UniBe).

Outline

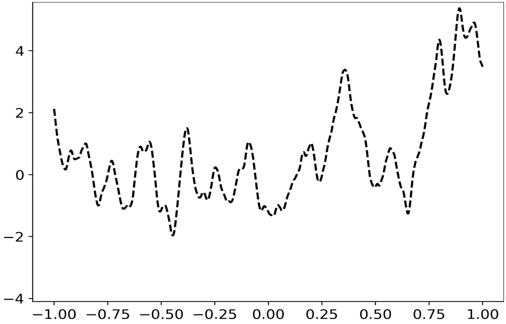
Section 1

Introduction

Problem Setup

Want to learn an unknown function

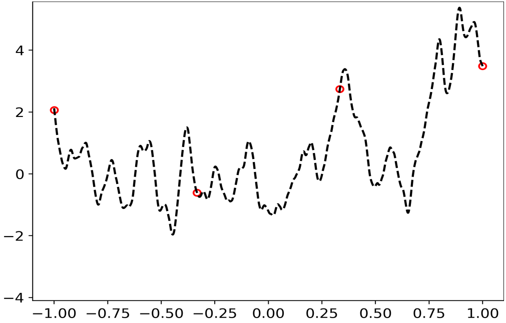
$$f : D \to \mathbb{R}$$



Problem Setup

Want to learn an unknown function

$$f : D \to \mathbb{R}$$



given some data $f(x_i)$.

Gaussian Process Regression

Can be done in a Bayesian way by assuming f is a realization of a Gaussian process prior $Z \sim \mathrm{Gp}(m_0, k)$.

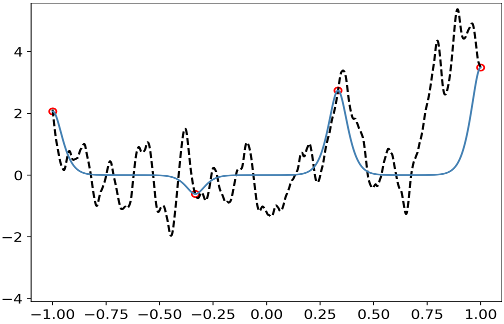


Figure: Posterior mean (blue).

Approximate f by posterior of Z conditional on the data $Z_{x_i} = f(x_i)$.

What if do not have point data $f(x_i)$ but more general data:

$$y_i = \ell_i(f)$$

for some linear functionals ℓ_i.

What if do not have point data $f(x_i)$ but more general data:

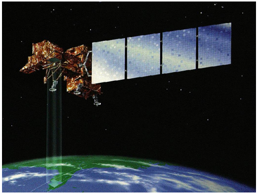
$$y_i = \ell_i(f)$$

for some linear functionals ℓ_i.

Examples

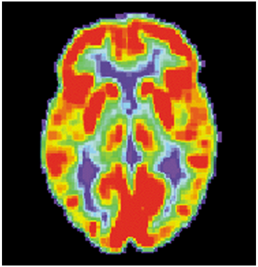
- Derivative observations $\ell_i(f) = f'(x_i)$
- Integral data: $\ell(f) = \int_D f(x)dx$
- Fourier coefficients $\ell_k(f) = \int_D e^{-2\pi ikx} f(x)dx$
- Kernel operators $\ell_s(f) = \int_D f(x)g(x,s)dx$, for some function g.

Linear operator data arise everywhere in science:



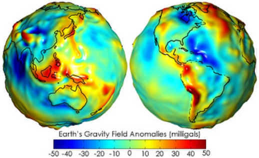
Remote Sensing

- Observe reflected light
- Recover land properties



Tomography

- Observe transmitted X-Ray intensity
- Recover material properties



Geoscience

- Observe gravitational field
- Recover undergroud properties

Broadly known as (linear) **Inverse Problems**.

Inverse Problems and GP

GPs can easily handle linear operator data.

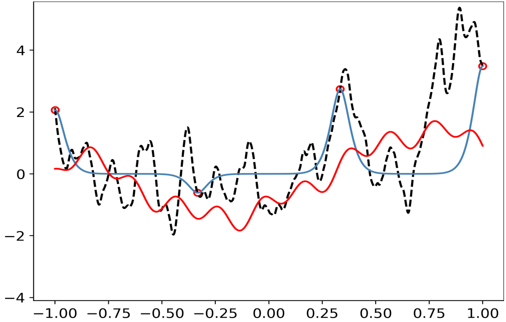


Figure: Posterior mean (red) for Fourier data

$$\ell_k(f) = \int_D e^{-2\pi i k x} f(x) dx, \; k = 1, 5, 7, 10$$

.

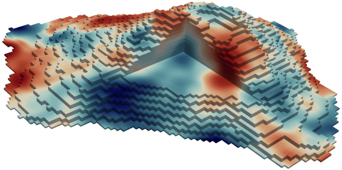
$\Longrightarrow$ can use GP priors in inverse problems (Bayesian Inversion)

Problem for today:

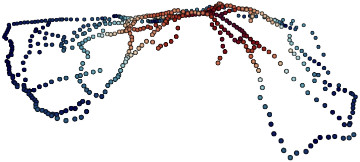
Can we scale the GP + linear operator framework to "real-world" problems?

Example Real-World Problem: Gravimetric Inversion

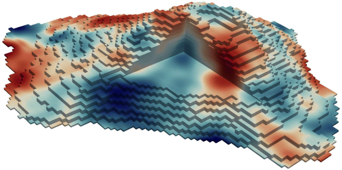
Recover interior of Stromboli volcano from surface gravity.



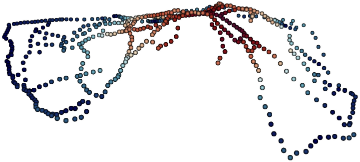
density



observed gravity



density



observed gravity

Properties:

- **Linear operator data**.
- **"Large-scale"**: large 3 dimensional inversion grid.
- **Sequential** assimilation of new data and **optimal design of experiment** important in practice.

Gravity field $G(s_i, \rho)$ at site s_i generated by underground density ρ can be written as linear operator:

$$y_i = G(s_i, \rho) = \int_D \rho(x) g(x, s_i) dx$$

Gravity field $G(s_i, \rho)$ at site s_i generated by underground density ρ can be written as linear operator:

$$y_i = G(s_i, \rho) = \int_D \rho(x) g(x, s_i) dx$$

- Traditionally solved by discretizing on a grid $\boldsymbol{x} = (x_1, ..., x_m)$.
- Observation model

$$\boldsymbol{y} = \boldsymbol{G}(\rho(x_1), ..., \rho(x_m))^T$$

.

- Discretized observation operator $\boldsymbol{G} \in \mathbb{R}^{n \times m}$ for n observations.
- Posterior described by:
  - mean vector $\tilde{\boldsymbol{m}} = (\tilde{m}_{x_1}, ..., \tilde{m}_{x_m})^T$
  - covariance matrix $\tilde{\boldsymbol{K}} = \left(\tilde{k}(x_i, x_j)\right)_{i,j=1,...,m}$

The Problem with Large, Sequential Bayesian Inversion

$$\tilde{\boldsymbol{m}} = \boldsymbol{m}_0 + \boldsymbol{K}\boldsymbol{G}^T \left(\boldsymbol{G}\boldsymbol{K}\boldsymbol{G}^T + \tau^2 \boldsymbol{I}\right)^{-1} (\boldsymbol{y} - \boldsymbol{G}\boldsymbol{m}_0)$$
$$\tilde{\boldsymbol{K}} = \boldsymbol{K} - \boldsymbol{K}\boldsymbol{G}^T \left(\boldsymbol{G}\boldsymbol{K}\boldsymbol{G}^T + \tau^2 \boldsymbol{I}\right)^{-1} \boldsymbol{G}\boldsymbol{K}$$

Linear operator data (can) involve **all** grid points at the same time.

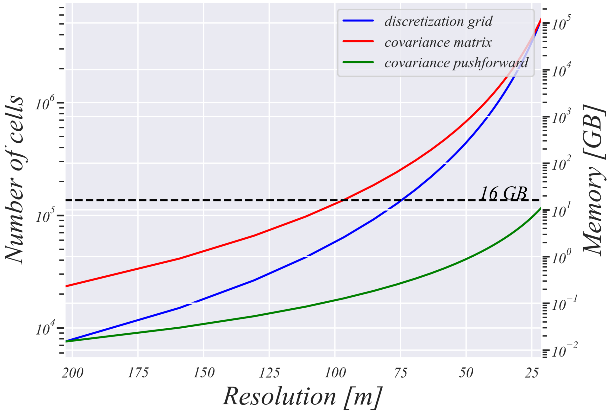


Figure: Grid and matrices size vs resolution on Stromboli example.

Section 2

Implicit Covariance Representation for Large-Scale Inversion

Solving **practical** difficulties of Bayesian inversion

- Covariance matrix too big? $\implies$ Don't store it, nor build it.

Implicit Representation

Posterior covariance information may be extracted via products with *tall and thin* matrices:

$$\tilde{K}A, \; A \in \mathbb{R}^{m \times p}, \; p \ll m$$

$\implies$ Only need to maintain a multiplication routine.

Travelletti, C., Ginsbourger, D. and Linde, N. (2022). *Uncertainty Quantification and Experimental Design for Large-Scale Linear Inverse Problems under Gaussian Process Priors* *https://arxiv.org/abs/2109.03457*, to appear in **SIAM JUQ**.

Implicit Representation: Advantages

- Drastically reduced memory footprint.

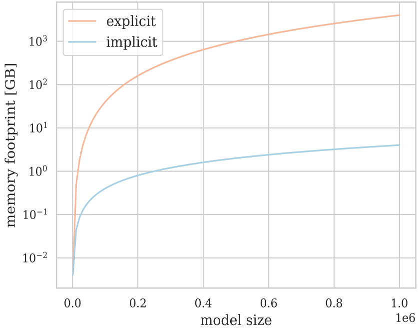


Figure: Memory footprint of posterior covariance vs grid size.

- Fast inclusion of new data.
- Update done in small chunks $\implies$ can send to GPU.

Inversion results (Matern 3/2 kernel), hyperparameters trained with MLE on field data are in agreement with deterministirc inversion results.

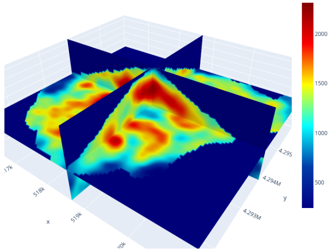


Figure: Posterior mean $[kg/m^3]$.

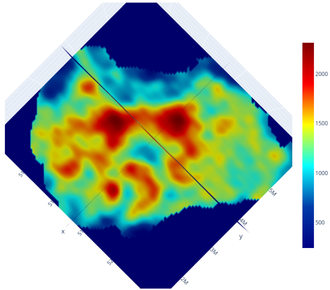


Figure: Posterior mean $[kg/m^3]$.

Probabilistic model allows for uncertainty quantification and active learning.

Theoretical Justification

Rigorous introduction of such an implicit representation requires us to understand which linear operators are allowed for the conditional law to be well defined.

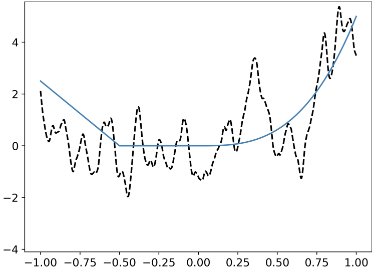
- Language of disintegrations of measures provides rigorous formulation of conditioning wrt. linear operators.
- Observation operator has to be a bounded operator into a separable Banach space.
- Need the GP to induce a Gaussian measure (valid in the usual settings).

Travelletti, C. and Ginsbourger, D. (2022). *Disintegration of Gaussian Measures for Sequential Bayesian Learning with Linear Operator Data* arxiv: 2207.13581 (2022), submitted to **EJS**.

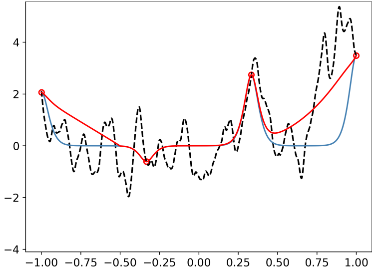
Section 3

Including known Trends

GP regression with trends known as **universal kriging**.

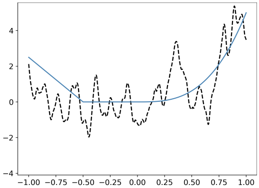


Ground truth

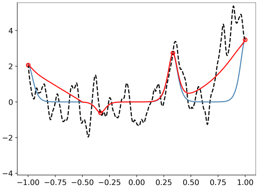


Point data

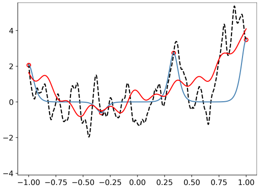
GP regression with trends known as **universal kriging**.



Ground truth



Point data



Fourier data

Goal: Extend to inverse problems to get **"universal inversion"**.

Trends allow inclusion of **expert knowledge**.

Model known geological stuctures such as:

- Layers
- Chimneys
- Depth-dependence
- ...

$$\text{target} = \text{trend} + \text{fluctuations}$$

Trends modeled by specifying basis functions:

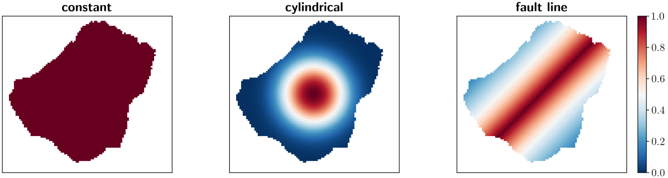


Figure: Basis functions (density at sea level $[kg/m^3]$).

Trends modeled by specifying basis functions:

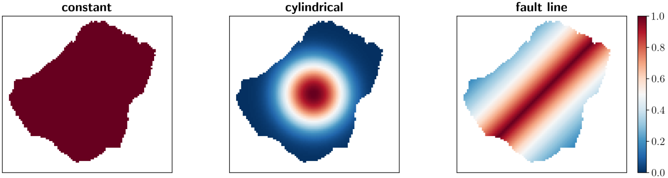


Figure: Basis functions (density at sea level $[kg/m^3]$).

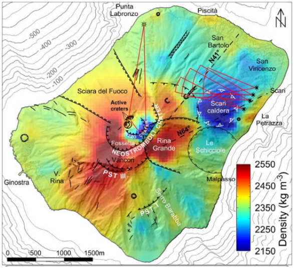


Figure: Structural boundaries of Stromboli Island [LBR+14].

Prior with trend

Assume prior is sum of trend + fluctuations described by Gaussian process

$$Z_x = \sum_{i=1}^{p} \beta_i f_i(x) + \eta_x,$$

- η is a (centred) GP with kernel k,
- basis functions $f_i : D \to \mathbb{R}$,
- trend coefficients $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$.

Put a Gaussian prior on the trend coefficients

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

Theorem

Conditionally on linear operator data $\boldsymbol{Y} = GZ_{\boldsymbol{w}} + \boldsymbol{\epsilon}$, *the posterior of the trend coefficients is Gaussian with mean and covariance given by:*

$$\begin{aligned}\mathbb{E}\left[\boldsymbol{\beta}|\boldsymbol{y}\right] &= \boldsymbol{\mu} + \Sigma\mathcal{F}_{\boldsymbol{w}}^T G^T Q_{\boldsymbol{y}}^{-1}\left(\boldsymbol{y} - G\mathcal{F}_{\boldsymbol{w}}\boldsymbol{\mu}\right)\\ \mathrm{Cov}\left(\boldsymbol{\beta}, \boldsymbol{\beta}|\boldsymbol{y}\right) &= \Sigma - \Sigma\mathcal{F}_{\boldsymbol{w}}^T G^T Q_{\boldsymbol{y}}^{-1} G\mathcal{F}_{\boldsymbol{w}}\Sigma,\end{aligned}$$

assuming that the matrix $Q_{\boldsymbol{y}} := G\left(\mathcal{F}_{\boldsymbol{w}}\Sigma\mathcal{F}_{\boldsymbol{w}}^T + K_{\boldsymbol{w}\boldsymbol{w}}\right)G^T + \sigma_\epsilon^2 I$ *is invertible.*

Conditionally on the data, the distribution of Z *is also that of a GP, with mean and covariance function given by:*

$$\begin{aligned}m_{Z|\boldsymbol{y}}\left(\boldsymbol{x}\right) &= \mathcal{F}_{\boldsymbol{x}}\boldsymbol{\mu} + \left(\mathcal{F}_{\boldsymbol{x}}\Sigma\mathcal{F}_{\boldsymbol{w}}^T + K_{\boldsymbol{x}\boldsymbol{w}}\right)G^T Q_{\boldsymbol{y}}^{-1}\left(\boldsymbol{y} - G\mathcal{F}_{\boldsymbol{w}}\boldsymbol{\mu}\right)\\ k_{Z|\boldsymbol{y}}\left(\boldsymbol{x}, \boldsymbol{x'}\right) &= K_{\boldsymbol{x}\boldsymbol{x'}} + \mathcal{F}_{\boldsymbol{x}}\Sigma\mathcal{F}_{\boldsymbol{x'}}^T - \left(\mathcal{F}_{\boldsymbol{x}}\Sigma\mathcal{F}_{\boldsymbol{w}}^T + K_{\boldsymbol{x}\boldsymbol{w}}\right)G^T Q_{\boldsymbol{y}}^{-1} G\\ &\qquad \left(\mathcal{F}_{\boldsymbol{w}}\Sigma\mathcal{F}_{\boldsymbol{x'}}^T + K_{\boldsymbol{w}\boldsymbol{x'}}\right)\end{aligned}$$

Volcano Inversion with Trend

Inversion with trends gives consistent results (data fit within noise std).

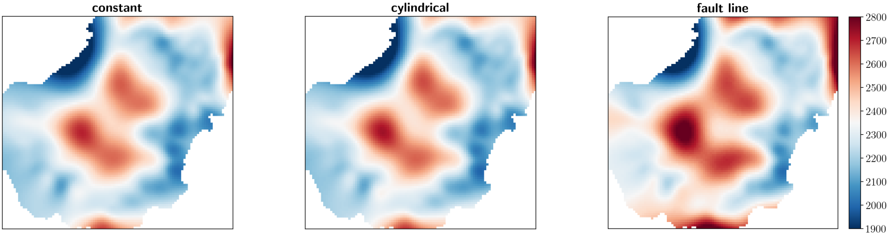


Figure: Posterior mean at -100 m $[kg/m^3]$

But gives rise to new questions.

Universal Inversion brings new challenges:

Questions

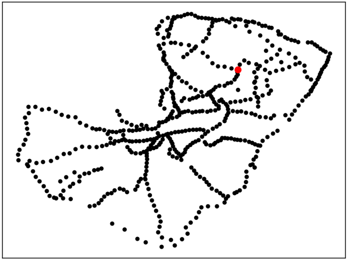
1. Can we compare the performance of different trend models?
2. How can we "rank" different trend models? (penalize for complexity)
3. Can we perform measurements to "optimally" discriminate models?

Cross-Validation

Idea

Given a data vector $\boldsymbol{Y} = GZ_{\boldsymbol{w}} + \boldsymbol{\epsilon}$:

1. Remove one data point Y_i.
2. Fit model on remaining data $\boldsymbol{Y}^{(-i)}$.
3. Use fitted model to predict missing data, compute residual $\hat{Y}_i^{(-i)} - Y_i$.



Residuals $\sum_{i=1}^{n} ||\hat{Y}_i^{(-i)} - Y_i||^2$ give measure of model generalisation error.

Cross-Validation (leave-k-out)

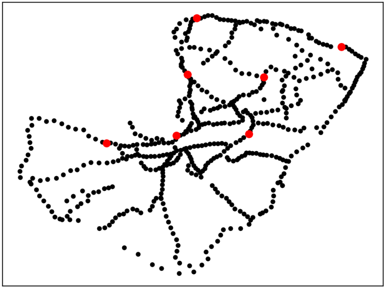
Can also consider subsets of size k: separate data in two batches:

$$\boldsymbol{Y} = (\boldsymbol{Y_i}, \boldsymbol{Y_{-i}}),$$

where $\boldsymbol{Y_i}$ has size k.

Want to compute residual when we predict $\boldsymbol{Y_i}$ using $\boldsymbol{Y_{-i}}$

$$\boldsymbol{Y_i} - \hat{\boldsymbol{Y}}_{\boldsymbol{i}}^{(-i)}$$



Problem

CV is **computationally expensive** (have to re-fit model at every pass).

Fast k-fold Cross-Validation Formulae

By generalizing [GS] all the information we need is contained in the augmented matrix:

$$\tilde{K} = \begin{pmatrix} GKG^T & GF \\ F^T G^T & \mathbf{0} \end{pmatrix}.$$

We partition it as:

$$\tilde{K} = \begin{pmatrix} \tilde{K}_{\boldsymbol{ii}} & \tilde{K}_{\boldsymbol{i-i}} \\ \tilde{K}_{\boldsymbol{-ii}} & \tilde{K}_{\boldsymbol{-i-i}} \end{pmatrix} = \begin{pmatrix} G_{\boldsymbol{i}\bullet} K G^T_{\bullet\boldsymbol{i}} & G_{\boldsymbol{i}\bullet} K G^T_{\bullet\boldsymbol{-i}} & G_{\boldsymbol{i}\bullet} F \\ G_{\boldsymbol{-i}\bullet} K G^T_{\bullet\boldsymbol{i}} & G_{\boldsymbol{-i}\bullet} K G^T_{\bullet\boldsymbol{-i}} & G_{\boldsymbol{-i}\bullet} F \\ F^T G^T_{\bullet\boldsymbol{i}} & F^T G_{\bullet\boldsymbol{-i}} & \mathbf{0} \end{pmatrix}.$$

CV residuals can the be computed by extracting subblocks of the inverse.

Fast k-fold Cross-Validation Formulae

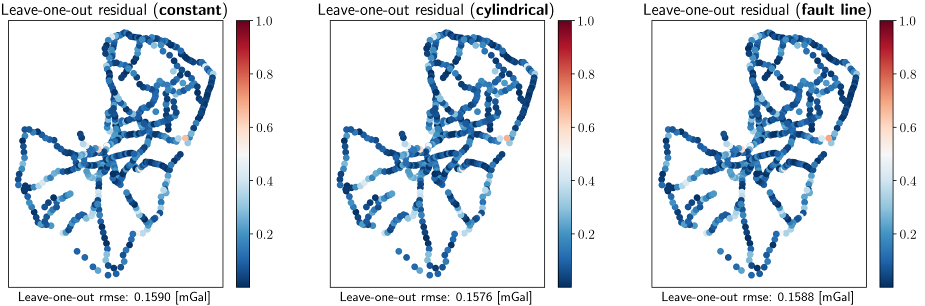
Upper left block of the inverse give us (inverse) covariance of residuals:

$$\tilde{K}_{\boldsymbol{ii}}^{-1} = \mathrm{Cov}\left(\hat{\boldsymbol{Y}}_{\boldsymbol{i}}^{(-\boldsymbol{i})}, \hat{\boldsymbol{Y}}_{\boldsymbol{i}}^{(-\boldsymbol{i})}\right)^{-1},$$

where $\tilde{K}_{\boldsymbol{ii}}^{-1}$ denotes the $\boldsymbol{ii}$ sub-block of $\tilde{K}^{-1}$. Can get the residuals in the same way:

$$\tilde{K}_{\boldsymbol{ii}}^{-1} \left(\tilde{K}^{-1}\left[:, 1:n\right]\boldsymbol{Y}\right)_{\boldsymbol{i}} = \boldsymbol{Y}_{\boldsymbol{i}} - \hat{\boldsymbol{Y}}_{\boldsymbol{i}}^{(-\boldsymbol{i})}.$$

Leave-One Out (LOO) Cross-Validation on Stromboli



Probems

- Gravimetric observations highly correlated $\Longrightarrow$ LOO not informative.
- Full leave-k-out infeasible (combinatorial explosion).

Need to select "representative" subsets.

K-fold Cross-Validation (experimental)

Instead of considering all data subsets of a given size, consider k different subsets (folds).

- Remove one fold and predict using remaining ones.

Choosing Folds

How can we choose the k subsets?

- Spatial clustering (heuristic).
- Data collection paths (heuristic).
- Other approach?

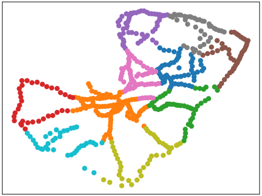
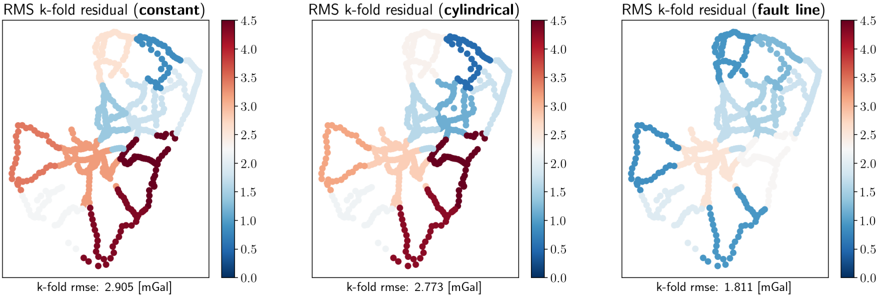


Figure: Folds from kMeans clustering, $k = 10$.

K-fold Cross-Validation: Stromboli (experimental)

kMean CV favors **fault line** model.



Question?

Is this representative of generalisation error?

K-fold Cross-Validation: Stromboli (experimental), contd.

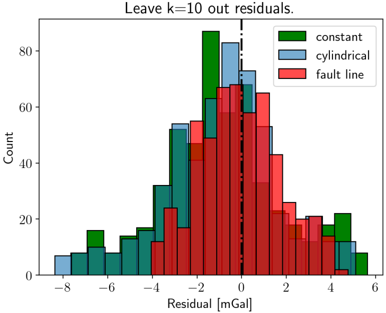


Figure: Histogram of k-fold residuals (aggregated across folds).

Open Questions and Future work

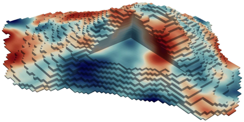
1. Principled model selection (penalize complexity).
2. Approach for choosing folds.
3. Active learning: choose next observations to optimally discriminate models.

Section 4

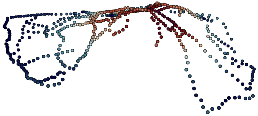
Optimal Design for Sequential Uncertainty Reduction on Excursion Sets

Excursion Set Estimation

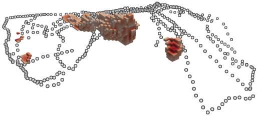
- Want to recover high-density regions.
- Correspond to geological features of interest.
- Propose *adaptive* data collection plan.



(a) density



(b) observation sites (static design)



(c) excursion set

Stepwise Uncertainty Reduction and Random Sets

Want to recover (true) excursion set $\Gamma_{\mathsf{true}} := \{x \in D : \ \rho(x) > T\}$.

Sequential Uncertainty Reduction and Random Sets

- Collect data $G_1, ..., G_n$.
- Compute posterior.
- Approximate through random set $\Gamma^{(n)} := \{x \in D : \ Z_x^{(n)} > T\}$, where $Z^{(n)}$ distributed according to posterior.

Stepwise Uncertainty Reduction and Random Sets

Want to recover (true) excursion set $\Gamma_{\text{true}} := \{x \in D : \; \rho(x) > T\}$.

Sequential Uncertainty Reduction and Random Sets

- Collect data $G_1, ..., G_n$.
- Compute posterior.
- Approximate through random set $\Gamma^{(n)} := \{x \in D : \; Z_x^{(n)} > T\}$, where $Z^{(n)}$ distributed according to posterior.

Coverage Function

$$
\begin{aligned}
p_n : & D \to [0,1] \\
p_n(x) := & \mathbb{P}\left[x \in \Gamma^{(n)}\right].
\end{aligned}
$$

Stepwise Uncertainty Reduction (SUR)

SUR Framework

Select observations by sequentially maximizing an uncertainty reduction criterion.

(weighted) IVR criterion

$$\mathrm{wIVR}^n(s) := \int_D \left(K_{xx}^{(n)} - K_{xx}^{(n+1)} \left[G_s \right] \right) p_n(x) dx$$

Where $K_{xx}^{(n+1)}$ denotes the conditional covariance after including a gravimetric observation $G_{\boldsymbol{s}}$ at location $\boldsymbol{s}$.

- Future variance $K^{(n+1)}$ 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*.

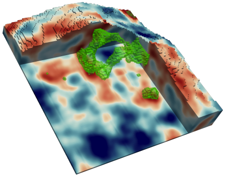


Figure: True density and excursion set (generated from model).

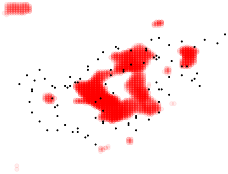


Figure: Proposed data collection plan, wIVR long-range strategy, total budget of 90 observations.

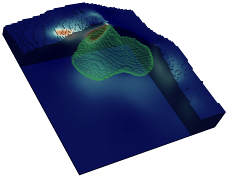


Figure: Estimated excursion set (Vorob'ev Expectation) and coverage function.

Results

Get data collection plan by myopic optimization of the criterion.

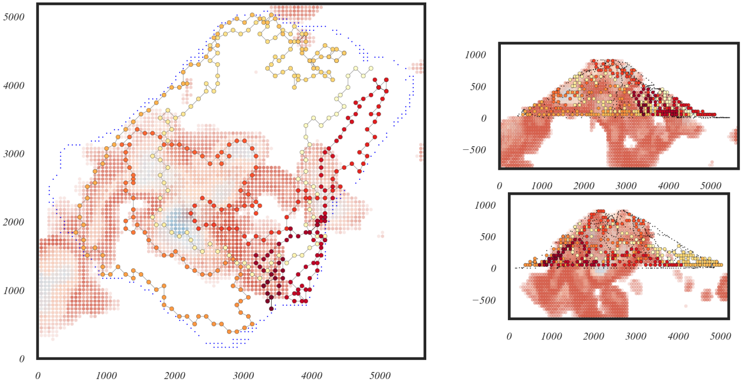
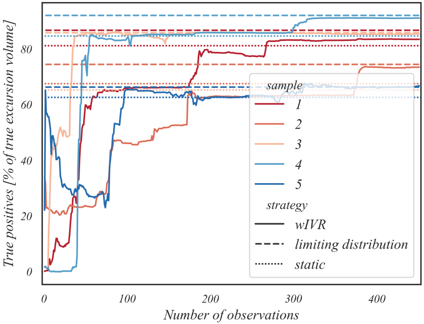


Figure: Excursion set and proposed data collection plan (tIMSE).

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

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

UQ on Excursion Volume

Sample from posterior volume distribution on 5 different ground truths.

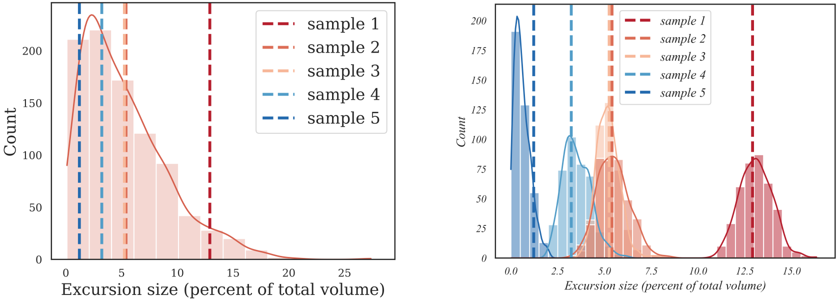


Figure: Prior (left) and empirical posterior (right) distribution (after 450 observations) of the excursion volume for each ground truth. True volumes are denoted by vertical lines.

Demonstrate uncertainty reduction and peaking around true volume.

Section 5

Towards more realistic Path Planning

Myopic optimization of **reward-only** criterion falls short of capturing reality.

Shortcomings

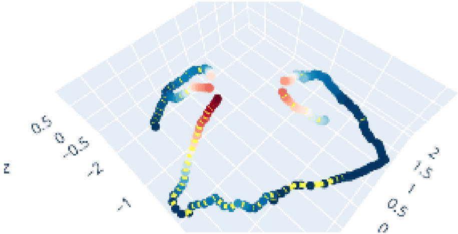
- Inacessible regions (Sciara del fuoco).
- Locations outside hiking trails difficult to reach.
- Some regions only accessible by boat.

And global constraint:

Have to be back to base at end of the day.

Cost Modelling

- Develop cost map of Stromboli (work in progress).
- How to compare cost (time spent) to reward (reduction of uncertainty).



Non-Myopic Optimization

Myopic optimization can get us "stuck" in bad regions.

$\implies$ Need to consider "lookahead criterion".

Possible Solution: Rollout

$$\mathrm{wIVR}^{n+k}(s) := \int_D \left(K_{xx}^{(n)} - K_{xx}^{(n+k)} \left[G_s \right] \right) p_{\boldsymbol{n}}(x) dx$$

Future variance independent of observations, future excursion probability approximated by current one.

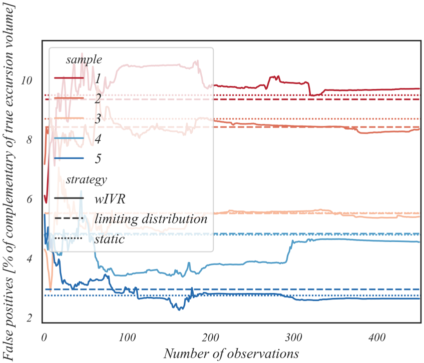
References I

- David Ginsbourger and Cedric Schärer, *Fast calculation of gaussian process multiple-fold cross-validation residuals and their covariances*.
- Niklas Linde, Ludovic Baron, Tullio Ricci, Anthony Finizola, André Revil, Filippo Muccini, Luca Cocchi, and Cosmo Carmisciano, *3-D density structure and geological evolution of Stromboli volcano (Aeolian Islands, Italy) inferred from land-based and sea-surface gravity data*, Journal of Volcanology and Geothermal Research **273** (2014), 58–69.

Image Sources

All images in the public domain.

- https://commons.wikimedia.org/wiki/File:Gravity_anomalies_on_Earth.jpg
- https://www.nasa.gov/audience/foreducators/robotics/imagegallery/r_landsat.jpg.html
- https://en.wikipedia.org/wiki/Brain_positron_emission_tomography#/media/File:PET_Normal_brain.jpg



(a) False positives