

Uncertainty Quantification on target regions under indirect measurements with applications to volcano geophysics

Cedric Travelletti

Idiap Research Institute & University of Bern

September 2, 2019

- 1 Problem Overview
- 2 Physical Setup
- 3 Bayesian Inversion
- 4 Hyperparameter Estimation
- 5 Model Selection
- 6 Set Estimation and Uncertainty Quantification on Sets
- 7 Next Steps

Section 1

Problem Overview

Problem Description

- Recovering the mass density field inside the Stromboli volcano from gravimetric measurements.
- Linear inverse problem, usually solved by least square + regularization.
- We here take a bayesian approach, using gaussian process priors.
- Allows recent advances in ML to be brought to geophysics
- The methods yields a posterior distribution over mass density fields that enables finer analysis than a deterministic solution.



<https://www.itij.com/story/115685/tourists-flee-stromboli-volcano-eruption>

Recovering mass density from gravimetric measurements

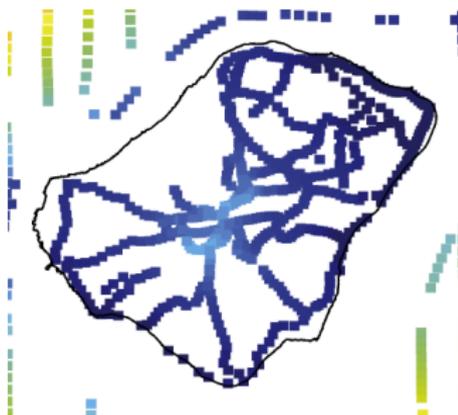
- Measure (relative) vertical component of the gravitational field at discrete locations on the surface of the volcano.
- Solve the inverse problem to reconstruct mass density inside.
- Test case on Stromboli island: 543 measurements, reconstruct density at 50m resolution ($\sim 200'000$ cells).
- Data provided by N. Linde's group (see Linde et al. 2014).

- Provide estimates of target regions (high density, high gradient, ...).
- UQ on target regions
- (future) Sequentially select new measurement locations for optimal reconstruction of the target regions.

Section 2

Physical Setup

- Let $u(x)$ denote the mass density field inside the volcano. We require it to be square integrable and to vanish on the boundary: $u \in L_0^2(X)$.
- We observe the z -component of the gravitational field at discrete locations $y_1, \dots, y_d \in \partial X$ on the boundary.



For a given location on the boundary $y_0 \in \partial X$, the z -component of the gravitational field generated by a mass distribution $u \in L^2_0(X)$ is given by:

$$g_z[u](y_0) = \gamma \int_X u(x) G_z(x, y_0) dx \quad (1)$$

Where γ is Newton's constant and we have the *Green function*:

$$G_z(x, y) = \frac{x_3 - y_3}{\|x - y\|^3}, \quad x = (x_1, x_2, x_3) \quad (2)$$

and the subscripts denote the components in the canonical basis of \mathbb{R}^3 .

Measurements at locations y_1, \dots, y_d on the boundary define a collection of linear forms $l_i := g_z[\cdot](y_i), i = 1, \dots, d$.

$$l_i : L_0^2(X) \rightarrow \mathbb{R}$$
$$u \mapsto \gamma \int_X u(x) G_z(x, y_i) d^3x$$

Collect the results of the measurements in a vector by defining:

$$\mathcal{F} : L_0^2(X) \rightarrow \mathbb{R}^d$$
$$u \mapsto \left(l_1(u), \dots, l_d(u) \right)^t$$

We call \mathcal{F} the *forward operator*.

Section 3

Bayesian Inversion

- Put a gaussian process prior on the mass density:
 $(Z_x)_{x \in X} \sim GP(\mu, k)$.
- Condition the field on the data $\ell_i[u] = g_i, i = 1, \dots, d$.
- Posterior distribution of the conditional field \tilde{Z}_x provides a distribution of *solutions* to the inverse problem.

The posterior distribution of the conditional field is a GP with covariance function and mean function given by

$$\begin{aligned}\tilde{\mu}(x) &= \mu(x) + l_i(k(x, \cdot)) \left(l_i(l_i(k(\cdot, \cdot))) + \Delta_\epsilon \right)^{-1} \left(g_i - l_i(\mu(\cdot)) \right) \\ \tilde{k}(x, x') &= k(x, x') - l_i(k(x, \cdot)) \left(l_i(l_i(k(\cdot, \cdot))) + \Delta_\epsilon \right)^{-1} l_i(k(x', \cdot))\end{aligned}$$

Inversion performed by updating the mean and covariance function.

Δ_ϵ denotes the noise variance.

- Inversion performed on a grid of 50 m. resolution.
- GP turns into a gaussian vector $Z_m := (Z_{x_1}, \dots, Z_{x_m})^t$.

Conditioning on $FZ_m = (d_1, \dots, d_d)^t =: d_{obs}$, posterior mean and covariance are

$$\begin{aligned}\tilde{\mu} &= \mu_0 + KF^T \left(FKF^T - \Delta_\epsilon \right)^{-1} (d_{obs} - F\mu_0) \\ \tilde{K} &= K - KF^T \left(FKF^T - \Delta_\epsilon \right)^{-1} FK\end{aligned}$$

Implementation Challenges related to the Conditioning

- Numerically, main challenge is to compute the $m \times d$ covariance pushforward KF^T , which involves the huge $m \times m$ covariance matrix.
- Due to the size of the covariance matrix ($4 * 10^{10}$ in our case), it can never be fully built, nor stored in memory. This also means that optimizing any parameter appearing inside it through gradient-based method is a lost cause.
- Matrix products computed on GPU after manual chunking.

Section 4

Hyperparameter Estimation

Maximum Likelihood

3 hyperparameters to optimize (in the squared exponential case): prior mean m_0 , prior variance σ_0^2 , lengthscale λ_0 .

- Estimate hyperparameters by MLE.
- m_0 by concentrating
- σ_0^2 by (automatic) gradient descent
- λ_0 by brute force grid search

The marginal data likelihood may be written as

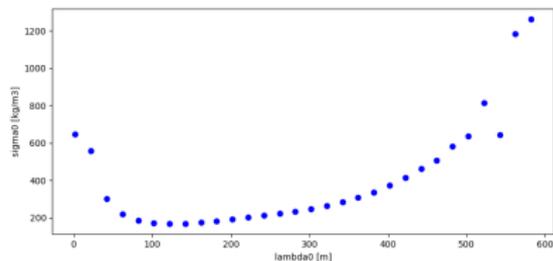
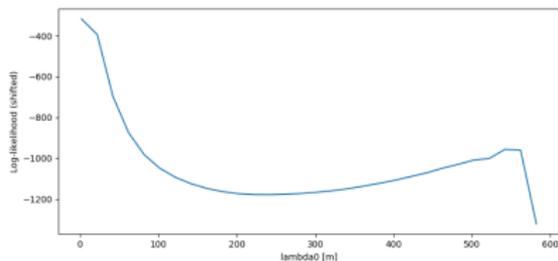
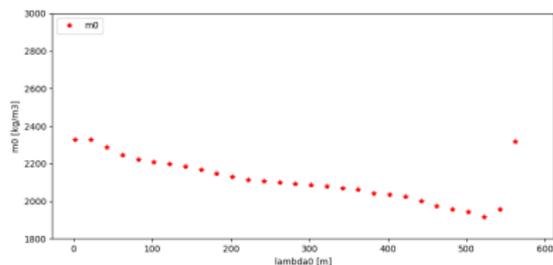
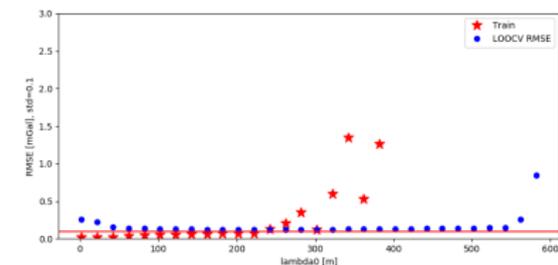
$$\begin{aligned} -2\mathcal{L}(\mu_0, \lambda_0, \sigma_0) &= n \log 2\pi - \log \det R^{-1} \\ &\quad + \left(d_{obs} - G\mu_0\right)^T R^{-1} \left(d_{obs} - G\vec{\mu}_0\right) \end{aligned}$$

Concentrating:

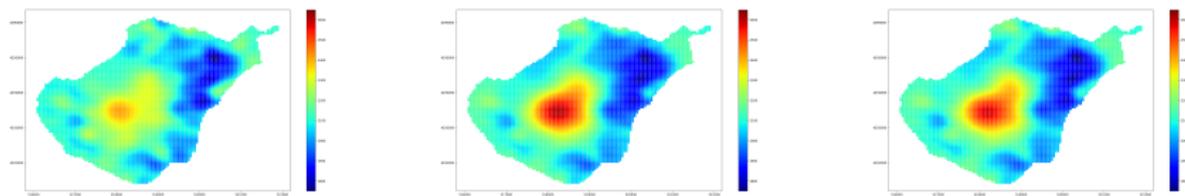
$$\hat{\mu}_0 = \left(I^T G^T R G I\right)^{-1} d_{obs}^T R G I$$

Hyperparameter Estimation

Optimization of the lengthscale parameter λ_0 for the squared exponential kernel.



Inversion Results



Sea level slice of posterior mean [mGal] for different kernels. From left to right: squared exponential, Matérn 3/2, Matérn 5/2.

Section 5

Model Selection

Cross-Validation

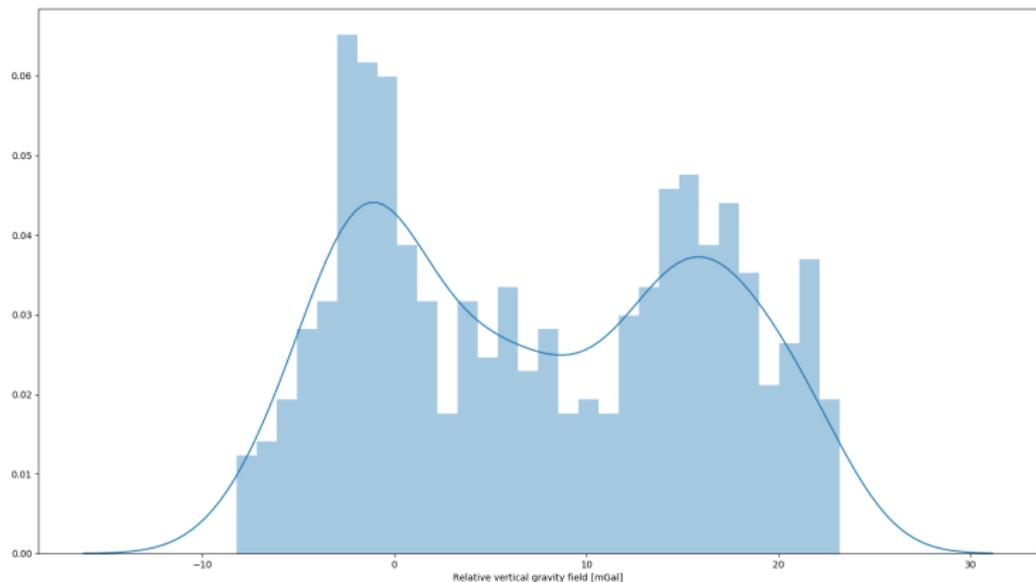
- Model selection by minimization of leave-one out root mean squared error.
- Remove one data point at a time and predict by conditioning on the remaining ones, average error over whole dataset.

Computation by brute force would be too expensive, fortunately, we have the fast-leave one out formula (adapted from Dubrule (1983)):

$$\hat{Z}_{x_{n+1}} = \mu_0(x_{n+1}) - \frac{1}{(n+1)\mathbf{R}_{n+1,n+1}^{-1}} \sum_{i=1}^n (n+1)\mathbf{R}_{n+1,i}^{-1} (y_i - \mu_0(x_i))$$

- Working on extending to k-fold cross-validation.

Data Distribution



Cross-Validation Results

Kernel	Train RMSE	LOOCV RMSE	Test RMSE
Squared exponential	0.073	0.119	0.147
Matérn 3/2	0.0735	0.114	0.138
Matérn 5/2	0.0755	0.116	0.136

Table: Root mean squared error [mGal].

Train/Test split: 350/192.

Section 6

Set Estimation and Uncertainty Quantification on Sets

We want to identify high density regions (excursion sets)

$$\Gamma^* = \{x \in X : u(x) \geq u_0\}$$

A simple plug-in estimate can be obtained using the posterior mean

$$\Gamma_{plug-in} = \{x \in X : \tilde{\mu}(x) \geq u_0\}.$$

Better estimates can be obtained by considering the full posterior distribution.

Random Closed Sets (RACS)

The posterior distribution of the conditional field gives rise to a random closed set (RACS) Γ

$$\Gamma = \{x \in X : \tilde{Z}_x \geq u_0\}$$

Can consider the pointwise probability to belong to the excursion set

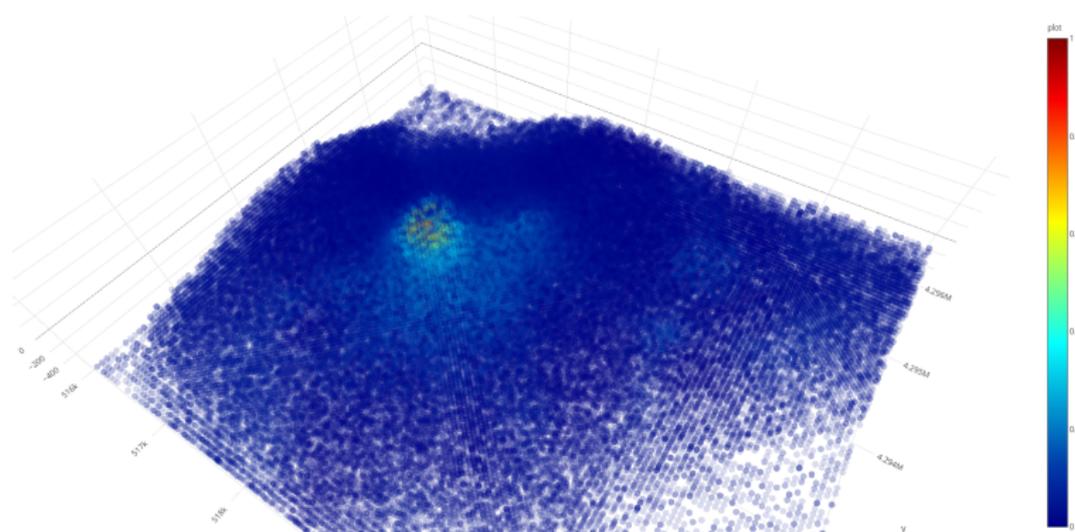
Coverage Function

$$p_\Gamma : X \rightarrow [0, 1]$$

$$p_\Gamma(x) := \mathbb{P}[x \in \Gamma]$$

Coverage function

Pointwise probability to belong the the excursion set above 2500 kg/m³.



The coverage function allows us to define a parametric family of set estimates for Γ

Vorob'ev Quantiles

$$Q_\alpha := \{x \in X : p_\Gamma \geq \alpha\}$$

The family of quantiles Q_α gives us a way to estimate Γ by controlling the (pointwise) probability α that the members of our estimate lie in Γ .

- Threshold α controls probability that points in our estimate lie in Γ .
- Can pick it such that the volume of the resulting set is equal to the expected volume of the excursion set

Vorob'ev Expectation

The Vorob'ev expectation is the quantile Q_{α_V} with threshold α_V chosen such that

$$\mu(Q_{\alpha_V}) = \mathbb{E}[\mu(\Gamma)]$$

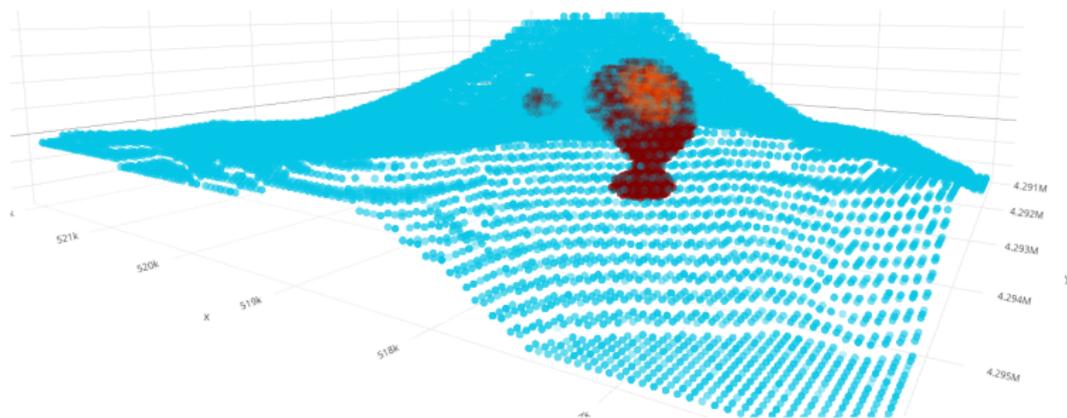
The expected volume of the excursion set can be computed using the coverage function

Robbins Theorem

$$\bar{V}_\Gamma := \mathbb{E}[\mu(\Gamma)] = \int_X p_\Gamma(x) dx$$

Vorob'ev Expectation

Plugin estimate and Vorob'ev expectation for excursion set above 2500.0 kg/m³.



Vorob'ev expectation: $\alpha = 0.22$, expected excursion measure $\mathbb{E}[\mu(\Gamma)] = 6678.16$ cells. Vorob'ev deviation: 7290.031 cells.

Can quantify uncertainty on an estimate Q for Γ by its Vorob'ev deviation

$$\mathcal{D}(Q) := \mathbb{E}[\mu(\Gamma \Delta Q)]$$

Theorem

$$\mathcal{D}(Q) = \int_Q (1 - p_\Gamma(x)) dx + \int_{Q^c} p_\Gamma(x) dx$$

This quantity will be the starting point for doing Bayesian optimal design, by selecting measurements that reduce the uncertainty.

Vorob'ev expectation achieves the minimum deviation among all sets that have measure equal to the expected measure of Γ .

Theorem

The Vorob'ev expectation minimizes the deviation among closed set with volume \bar{V}_Γ .

$$Q_{\alpha_V} \in \arg \min \{ \mathcal{D}(Q) \mid Q \subset X \text{ closed, } \mu(Q) = \bar{V}_\Gamma \}$$

Section 7

Next Steps

- New kernels (anisotropic, non-stationary).
- New random field models (gauss-markov random fields, ...).
- Sequential experimental design: select new measurement locations to improve estimate on a given target region.

-  D. Azzimonti, J. Bect, C. Chevalier, D. Ginsbourger: *Quantifying Uncertainties on Excursion Sets Under a Gaussian Random Field Prior*, SIAM/ASA J. Uncertainty Quantification, 4(1), 850874.
-  C. Chevalier, D. Ginsbourger, J. Bect, I. Molchanov: *Estimating and Quantifying Uncertainties on Level Sets Using the Vorobev Expectation and Deviation with Gaussian Process Models*, mODa 10 Advances in Model-Oriented Design and Analysis.
-  N. Linde, L. Baron, T. Ricci, A. Finizola, A. Revil, F. Muccini, L. Cocchi, C. 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, Elsevier, 2014, 273, pp.58-69.
-  I. Molchanov: *Theory of Random Sets*, Springer, 2005.