Quantifying and reducing uncertainties on sets under Gaussian Process priors

David Ginsbourger [1,2]

Acknowledgements: a number of co-authors, notably appearing via citations!

[1] Idiap Research Institute, UQOD group, Martigny, Switzerland, and
[2] Department of Mathematics and Statistics, IMSV, University of Bern

Gaussian Process and Uncertainty Quantification Summer School

September 2018, University of Sheffield

Preamble

Set up: estimate a deterministic function $f : \mathbf{x} \in E \mapsto f(\mathbf{x}) \in F$ and/or quantities relying on it based on a limited number of evaluations of f.

Preamble

Set up: estimate a deterministic function $f : \mathbf{x} \in E \mapsto f(\mathbf{x}) \in F$ and/or quantities relying on it based on a limited number of evaluations of f.

Two typical examples where f stems from numerical simulations

- Safety engineering: $\mathbf{x}$ is a vector parametrizing some system and f returns an indicator of dangerousness. It is then crucial to understand which $\mathbf{x}$'s lead to "high" values of $f(\mathbf{x})$.

Preamble

Set up: estimate a deterministic function $f : \mathbf{x} \in E \mapsto f(\mathbf{x}) \in F$ and/or quantities relying on it based on a limited number of evaluations of f.

Two typical examples where f stems from numerical simulations

- Safety engineering: $\mathbf{x}$ is a vector parametrizing some system and f returns an indicator of dangerousness. It is then crucial to understand which $\mathbf{x}$'s lead to “high” values of $f(\mathbf{x})$.

- Flow simulation: $\mathbf{x}$ stands e.g. for the medium, boundary conditions, etc. and f returns the evolution of a fluid and/or a measure of discrepancy between simulation results and given observation results.

Preamble: Bayesian approach with GP models

Typical situation : f was evaluated at a set of “points” $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D \subset E$ and one wishes to estimate a quantity relying on f and/or run new evaluations in order to improve this estimation.

Preamble: Bayesian approach with GP models

Typical situation : f was evaluated at a set of “points” $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D \subset E$ and one wishes to estimate a quantity relying on f and/or run new evaluations in order to improve this estimation.

$\Rightarrow$ legitimate to rely on some approximation(s) of f knowing $f(\mathbf{x}_i) + \epsilon_i$ $(1 \leq i \leq n)$. A number of approaches do exist...

Preamble: Bayesian approach with GP models

Typical situation : f was evaluated at a set of “points” $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D \subset E$ and one wishes to estimate a quantity relying on f and/or run new evaluations in order to improve this estimation.

$\Rightarrow$ legitimate to rely on some approximation(s) of f knowing $f(\mathbf{x}_i) + \epsilon_i$ $(1 \leq i \leq n)$. A number of approaches do exist...

Principles of the Gaussian Process approach (GP): suppose that, *a priori*, f is a realization of a GP $(Z_{\mathbf{x}})_{\mathbf{x} \in D}$ and approximate f and/or the quantities of interest via the **conditional distribution** of Z knowing $Z_{\mathbf{x}_i} + \varepsilon_i = f(\mathbf{x}_i) + \epsilon_i$.

Preamble: Bayesian approach with GP models

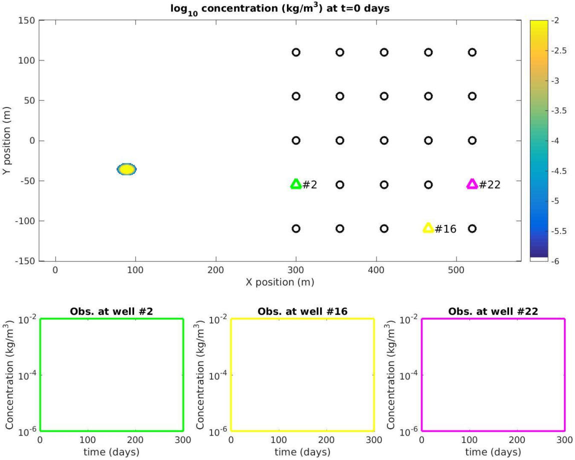
Typical situation : f was evaluated at a set of "points" $\mathbf{x}_1, \ldots, \mathbf{x}_n \in D \subset E$ and one wishes to estimate a quantity relying on f and/or run new evaluations in order to improve this estimation.

$\Rightarrow$ legitimate to rely on some approximation(s) of f knowing $f(\mathbf{x}_i) + \epsilon_i$ $(1 \leq i \leq n)$. A number of approaches do exist...

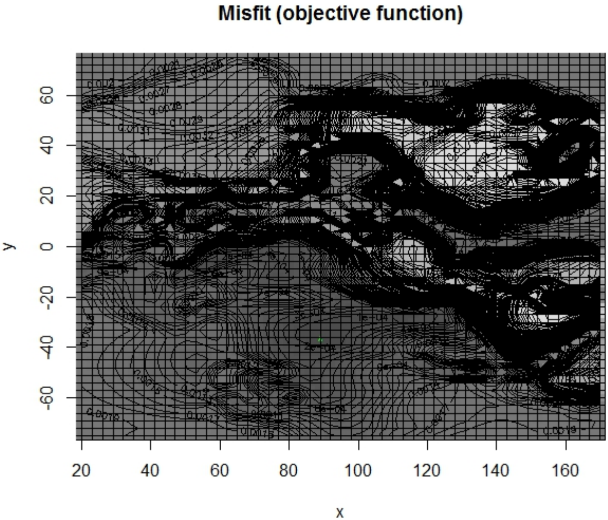
Principles of the Gaussian Process approach (GP): suppose that, *a priori*, f is a realization of a GP $(Z_{\mathbf{x}})_{\mathbf{x} \in D}$ and approximate f and/or the quantities of interest via the **conditional distribution** of Z knowing $Z_{\mathbf{x}_i} + \varepsilon_i = f(\mathbf{x}_i) + \epsilon_i$.

$\Rightarrow$ very practical for sequential design of experiments.

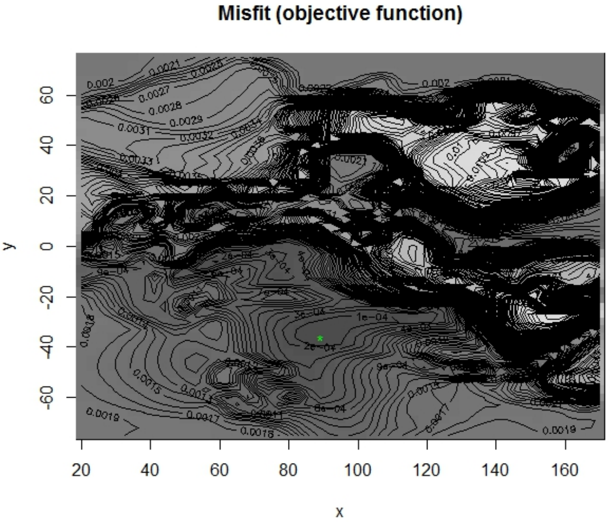
Preamble: example inverse problem in hydrogeology



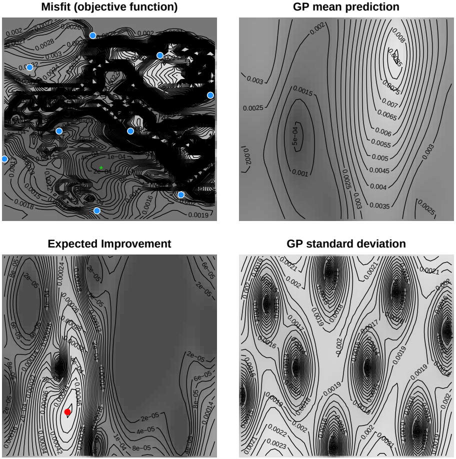
Preamble: a costly full factorial experimental design!



Preamble: a costly full factorial experimental design!



Preamble: an application of Bayesian optimization



The previous example was produced in the framework of an ongoing collaboration with G. Pirot (University of Lausanne), T. Krityakierne (now at Mahidol University, Bangkok) and P. Renard (University of Neuchâtel).

$\Rightarrow$ See ongoing Hydrol. Earth Syst. Sci. Discuss. paper (2017+).

The previous example was produced in the framework of an ongoing collaboration with G. Pirot (University of Lausanne), T. Krityakierne (now at Mahidol University, Bangkok) and P. Renard (University of Neuchâtel).

$\Rightarrow$ See ongoing Hydrol. Earth Syst. Sci. Discuss. paper (2017+).

Main focus today

In a related set-up, how to estimate excursion sets of f using such models and dedicated sequential design strategies?

The previous example was produced in the framework of an ongoing collaboration with G. Pirot (University of Lausanne), T. Krityakierne (now at Mahidol University, Bangkok) and P. Renard (University of Neuchâtel).

$\Rightarrow$ See ongoing Hydrol. Earth Syst. Sci. Discuss. paper (2017+).

Main focus today

In a related set-up, how to estimate excursion sets of f using such models and dedicated sequential design strategies?

As a transition, let us review a few selected seminal references about GP modelling and GP-based "Bayesian" optimization.

A few references on GP modelling . . .

A. O'Hagan (1978).
Curve fitting and optimal design for prediction.
Journal of the Royal Statistical Society, Series B, 40(1):1-42.

J. Sacks, W.J. Welch, T.J. Mitchell, and H. P. Wynn (1989).
Design and Analysis of Computer Experiments
Statist. Sci. 4(4), 409-423.

H. Omre and K. Halvorsen (1989).
The bayesian bridge between simple and universal kriging.
Mathematical Geology, 22 (7):767-786.

M. S. Handcock and M. L. Stein (1993).
A bayesian analysis of kriging.
Technometrics, 35(4):403-410.

A.W. Van der Vaart and J. H. Van Zanten (2008).
Rates of contraction of posterior distributions based on Gaussian process priors.
Annals of Statistics, 36:1435-1463.

. . . and on GP-based Optimization

H.J. Kushner (1964).
A new method of locating the maximum of an arbitrary multi-peak curve in the presence of noise.
Journal of Basic Engineering, 86:97-106.

J. Mockus (1972).
On Bayesian methods for seeking the extremum.
Automatics and Computers (Avtomatika i Vychislitel'naya Tekhnika), 4(1):53-62.

J. Mockus, V. Tiesis, and A. Zilinskas (1978).
The application of Bayesian methods for seeking the extremum.
In Dixon, L. C. W. and Szegö, G. P., editors, Towards Global Optimisation, volume 2, pages 117-129. Elsevier Science Ltd., North Holland, Amsterdam.

J.M. Calvin (1997).
Average performance of a class of adaptive algorithms for global optimization.
The Annals of Applied Probability, 7(3):711-730.

M. Schonlau, W.J. Welch and D.R. Jones (1998).
Efficient Global Optimization of Expensive Black-box Functions.
Journal of Global Optimization.

Our main topic today: background and motivations

A number of practical problems boil down to determining sets of the form

$$\Gamma^{\star} = \{\mathbf{x} \in D : f(\mathbf{x}) \in T\} = f^{-1}(T)$$

where $f : D \longrightarrow \mathbb{R}^k$ ($k \geq 1$), $D \subset \mathbb{R}^d$ ($d \geq 1$), and $T \subset \mathbb{R}^k$.

Our main topic today: background and motivations

A number of practical problems boil down to determining sets of the form

$$\Gamma^{\star} = \{\mathbf{x} \in D : f(\mathbf{x}) \in T\} = f^{-1}(T)$$

where $f : D \longrightarrow \mathbb{R}^k$ ($k \geq 1$), $D \subset \mathbb{R}^d$ ($d \geq 1$), and $T \subset \mathbb{R}^k$.

Examples

- Contour lines
- Excursion/sojourn sets above/below thresholds
- Admissible regions in constrained optimization
- High gradient/high curvature regions, etc.
- (Pareto sets in multi-objective optimization...but then T depends on f!)

Background and motivations

We essentially focus today on the case where $k = 1$, D is compact, f is continuous, and $T = [t, +\infty)$ or $(-\infty, t]$ for some prescribed $t \in \mathbb{R}$.

$\Gamma^\star = \{\mathbf{x} \in D : f(\mathbf{x}) \geq t\}$ is then referred to as the excursion set of f above t.

Background and motivations

We essentially focus today on the case where $k = 1$, D is compact, f is continuous, and $T = [t, +\infty)$ or $(-\infty, t]$ for some prescribed $t \in \mathbb{R}$.

$\Gamma^\star = \{\mathbf{x} \in D : f(\mathbf{x}) \geq t\}$ is then referred to as the excursion set of f above t.

Our aim is to estimate $\Gamma^\star$ and quantify uncertainty on it when f can solely be evaluated at a few points, both in static and sequential cases.

Test case from safety engineering

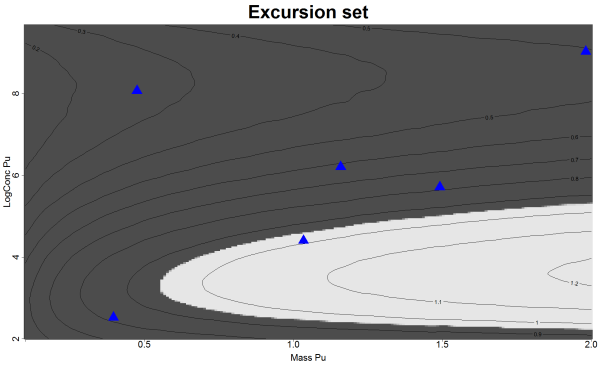


Figure: Excursion set (light gray) of a nuclear criticality safety coefficient depending on two design parameters. Blue triangles: initial experiments.

C. Chevalier (2013).
Fast uncertainty reduction strategies relying on Gaussian process models.
Ph.D. thesis, University of Bern.

Making a sensible estimation of $\Gamma^\star$ based on a drastically limited number of evaluations $f(\mathbf{X}_n) = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))'$ calls for additional assumptions on f.

Making a sensible estimation of $\Gamma^\star$ based on a drastically limited number of evaluations $f(\mathbf{X}_n) = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))'$ calls for additional assumptions on f.

As before, we consider the Bayesian framework where a Gaussian Process (GP) prior is put on f, i.e. f is seen as one realization of a GP $(Z(\mathbf{x}))_{\mathbf{x}\in D}$ (characterized in distribution by a mean m and a covariance kernel k).

Making a sensible estimation of $\Gamma^\star$ based on a drastically limited number of evaluations $f(\mathbf{X}_n) = (f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n))'$ calls for additional assumptions on f.

As before, we consider the Bayesian framework where a Gaussian Process (GP) prior is put on f, i.e. f is seen as one realization of a GP $(Z(\mathbf{x}))_{\mathbf{x}\in D}$ (characterized in distribution by a mean m and a covariance kernel k).

In the GP set-up, the main object of interest is represented by

$$\Gamma = \{\mathbf{x} \in D : Z(\mathbf{x}) \in T\} = Z^{-1}(T)$$

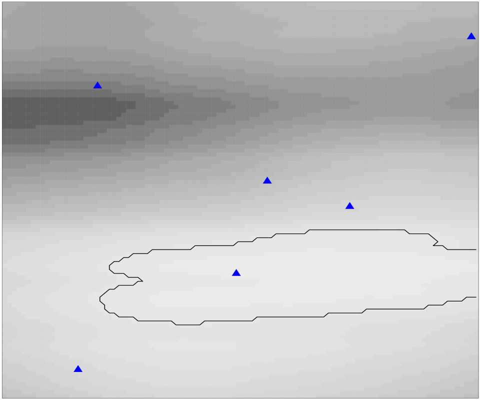
Under our previous assumptions on T and assuming that is chosen Z with continuous paths, Γ is a Random Closed Set (See thesis below for detail).

D. Azzimonti (2016).
Contributions to Bayesian set estimation relying on random field priors.
Ph.D. thesis, University of Bern.

Simulating excursion sets under a GRF model

Posterior simulations on a 50×50 grid of Z and Γ knowing $Z(\mathbf{X}_n) = f(\mathbf{X}_n)$.

Posterior GRF realization



How to quantify the uncertainty on Γ?

There are many ways to quantify uncertainties on sets!

How to quantify the uncertainty on Γ?

There are many ways to quantify uncertainties on sets!

This will be one of the recurring questions throughout the talk, but we will not be exhaustive by far. For more detail see, e.g.,

I. Molchanov (2005)
Theory of Random Sets.
Springer.

D. Azzimonti, J. Bect, C. Chevalier and D. Ginsbourger (2016).
Quantifying uncertainties on excursion sets under a Gaussian random field prior.
SIAM/ASA Journal on Uncertainty Quantification.

How to quantify the uncertainty on Γ?

There are many ways to quantify uncertainties on sets!

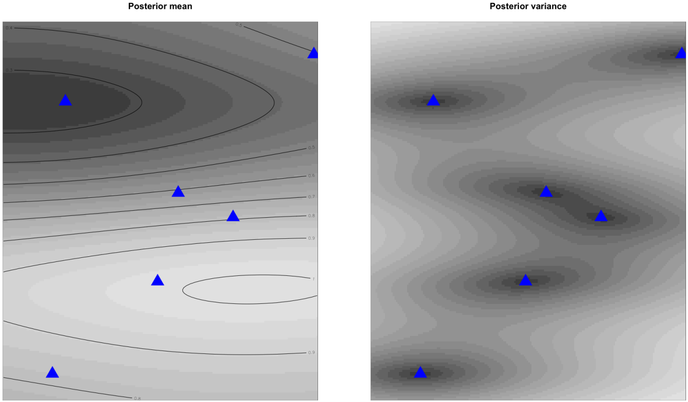
This will be one of the recurring questions throughout the talk, but we will not be exhaustive by far. For more detail see, e.g.,

I. Molchanov (2005)
Theory of Random Sets.
Springer.

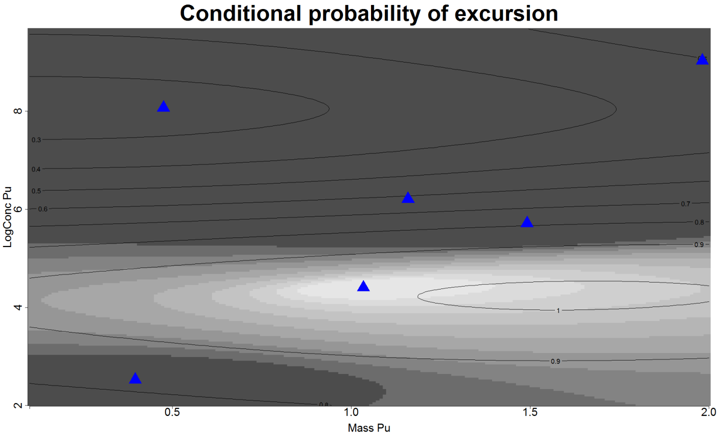
D. Azzimonti, J. Bect, C. Chevalier and D. Ginsbourger (2016).
Quantifying uncertainties on excursion sets under a Gaussian random field prior.
SIAM/ASA Journal on Uncertainty Quantification.

Before moving to random set-related concepts, a first spontaneous idea is to "scalarize" the problem, for instance by looking at Γ's volume. Let us make a detour through some GP basics in order to do so.

Kriging (Gaussian Process Interpolation)



$$\begin{cases} m_n(\mathbf{x}) = m(\mathbf{x}) + k(\mathbf{X}_n, \mathbf{x})^T k(\mathbf{X}_n, \mathbf{X}_n)^{-1} (f(\mathbf{X}_n) - m(\mathbf{X}_n)) \\ s_n^2(\mathbf{x}) = k(\mathbf{x}, \mathbf{x}) - k(\mathbf{X}_n, \mathbf{x})^T k(\mathbf{X}_n, \mathbf{X}_n)^{-1} k(\mathbf{X}_n, \mathbf{x}) \end{cases}$$



From $\mathcal{L}_n(Z_{\mathbf{x}}) = \mathcal{N}(m_n(\mathbf{x}), s_n^2(\mathbf{x}))$, the "coverage probability" of Γ (or conditional/posterior probability of excursion, here) can be expanded as

$$p_n(\mathbf{x}) = \mathbb{P}_n(\mathbf{x} \in \Gamma) = \mathbb{P}_n(Z(\mathbf{x}) \geq t) = \Phi\left(\frac{m_n(\mathbf{x}) - t}{s_n(\mathbf{x})}\right)$$

From p_n to moments of Γ's volume

Denote by μ a finite measure on $(D, \mathcal{B}(D))$ and set $\alpha^* = \mu(\Gamma^*)$, i.e. the "volume of excursion" in the considered case.

The GP model leads to a random analogue $\alpha = \mu(\Gamma)$, and by Robbins' theorem, the posterior expectation of α can be written in terms of p_n:

$$\mathbb{E}_n[\mu(\Gamma)] = \mathbb{E}_n\left[\int_D \mathbf{1}_\Gamma(\mathbf{u}) d\mu(\mathbf{u})\right] = \int_D p_n(\mathbf{u}) d\mu(\mathbf{u})$$

From p_n to moments of Γ's volume

Denote by μ a finite measure on $(D, \mathcal{B}(D))$ and set $\alpha^* = \mu(\Gamma^*)$, i.e. the "volume of excursion" in the considered case.

The GP model leads to a random analogue $\alpha = \mu(\Gamma)$, and by Robbins' theorem, the posterior expectation of α can be written in terms of p_n:

$$\mathbb{E}_n[\mu(\Gamma)] = \mathbb{E}_n\left[\int_D \mathbf{1}_\Gamma(\mathbf{u})d\mu(\mathbf{u})\right] = \int_D p_n(\mathbf{u})d\mu(\mathbf{u})$$

However, the (posterior) distribution of α has been considered analytically intractable.

R.J. Adler (2000)
On excursion sets, tube formulas and maxima of random fields.
Annals of Applied Probability, 10(1):1-74.

E. Vazquez and M. Piera Martinez (2006).
Estimation of the volume of an excursion set of a Gaussian process using intrinsic Kriging.
arXiv:math/0611273 [math.ST].

About conditional moments of α

Fortunately, as already pointed out in Molchanov 2005 in more general settings, $\mathbb{E}_n[\alpha^r]$ can also be worked out for $r \geq 2$), at the price of calculating integrals. In our framework, we have indeed:

$$
\begin{aligned}
\mathbb{E}_n[\alpha^r] &= \mathbb{E}_n\left[\left(\int_D \mathbf{1}_\Gamma(\mathbf{u}) d\mu(\mathbf{u})\right)^r\right] \\
&= \mathbb{E}_n\left[\left(\int_D \mathbf{1}_\Gamma(\mathbf{u}_1) d\mu(\mathbf{u}_1)\right) \dots \left(\int_D \mathbf{1}_\Gamma(\mathbf{u}_r) d\mu(\mathbf{u}_r)\right)\right] \\
&= \int_D \cdots \int_D \mathbb{E}_n\left[\mathbf{1}_\Gamma(\mathbf{u}_1) \dots \mathbf{1}_\Gamma(\mathbf{u}_r)\right] d\mu(\mathbf{u}_1) \dots d\mu(\mathbf{u}_r) \\
&= \int_D \cdots \int_D \mathbb{P}_n(Z_{\mathbf{u}_1} \geq t, \dots, Z_{\mathbf{u}_r} \geq t) d\mu(\mathbf{u}_1) \dots d\mu(\mathbf{u}_r)
\end{aligned}
$$

Hence, recalling the GP assumption, $\mathbb{E}_n[\alpha^r]$ writes as an r-dimensional integral which integrand involves a r-dimensional Gaussian CDF.

A useful bound for the case $r = 2$

In what follows, the case $r = 2$ will be of special importance as we will consider sequential design strategies aiming at reducing $\mathrm{Var}_n[\alpha]$.

The following underlined quantity, that is easier to compute and also comes with a nice interpretation, has been used as well:

$$
\begin{aligned}
\mathrm{Var}_n[\alpha] &= \mathbb{E}_n\left[\left(\int_D (\mathbf{1}_\Gamma(\mathbf{u}) - p_n(\boldsymbol{u}))d\mu(\boldsymbol{u})\right)^2\right] \\
&\leq \mu(D)^2 \mathbb{E}_n\left[\int_D (\mathbf{1}_\Gamma(\mathbf{u}) - p_n(\boldsymbol{u}))^2 d\mu(\boldsymbol{u})\right] \\
&= \mu(D)^2 \underbrace{\int_D p_n(\boldsymbol{u})(1 - p_n(\boldsymbol{u}))d\mu(\boldsymbol{u})}_{\text{Integrated indicator variance}}
\end{aligned}
$$

A useful bound for the case $r = 2$

In what follows, the case $r = 2$ will be of special importance as we will consider sequential design strategies aiming at reducing $\mathrm{Var}_n[\alpha]$.

The following underlined quantity, that is easier to compute and also comes with a nice interpretation, has been used as well:

$$\begin{aligned}
\mathrm{Var}_n[\alpha] &= \mathbb{E}_n\left[\left(\int_D (\mathbf{1}_\Gamma(\mathbf{u}) - p_n(\boldsymbol{u}))d\mu(\boldsymbol{u})\right)^2\right] \\
&\leq \mu(D)^2 \mathbb{E}_n\left[\int_D (\mathbf{1}_\Gamma(\mathbf{u}) - p_n(\boldsymbol{u}))^2 d\mu(\boldsymbol{u})\right] \\
&= \mu(D)^2 \underbrace{\int_D p_n(\boldsymbol{u})(1 - p_n(\boldsymbol{u}))d\mu(\boldsymbol{u})}_{\text{Integrated indicator variance}}
\end{aligned}$$

The excursion volume's variance and the integrated indicator variance are used as two particular “measures of uncertainty” in what follows.

Towards Stepwise Uncertainty Reduction strategies

Let us informally consider the following 1-step-lookahead scheme:

- For some chosen (say, non-negative) functional defined on GP distributions, define the uncertainty at time $n \geq 0$, H_n, as this functional applied to the current posterior GP (E.g., $H_n = var_n(\alpha)$).

- Starting from some intial design $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_0}\}$, at each iteration $n \geq n_0$, evaluate f at a point $\mathbf{x}^{\star}_{n+1}$ minimizing the so-called SUR criterion associated with the chosen notion of uncertainty:

$$J_n(\mathbf{x}_{n+1}) := \mathbb{E}_n(H_{n+1}(\mathbf{x}_{n+1}))$$

Towards Stepwise Uncertainty Reduction strategies

Let us informally consider the following 1-step-lookahead scheme:

- For some chosen (say, non-negative) functional defined on GP distributions, define the uncertainty at time $n \geq 0$, H_n, as this functional applied to the current posterior GP (E.g., $H_n = var_n(\alpha)$).

- Starting from some intial design $\{\mathbf{x}_1, \ldots, \mathbf{x}_{n_0}\}$, at each iteration $n \geq n_0$, evaluate f at a point $\mathbf{x}^\star_{n+1}$ minimizing the so-called SUR criterion associated with the chosen notion of uncertainty:

$$J_n(\mathbf{x}_{n+1}) := \mathbb{E}_n(H_{n+1}(\mathbf{x}_{n+1}))$$

See notably the following paper and seminal references therein:

J. Bect, D. Ginsbourger, L. Li, V. Picheny and E. Vazquez.
Sequential design of computer experiments for the estimation of a probability of failure.
Statistics and Computing, 22(3):773-793, 2012.

SUR strategies: Two candidate uncertainties

Two possible definitions for the uncertainty H_n are considered below:

$$
\begin{aligned}
H_n &:= Var_n(\alpha) \\
\widetilde{H}_n &:= \int_D p_n(1 - p_n)\mathrm{d}\mu
\end{aligned}
$$

SUR strategies: Two candidate uncertainties

Two possible definitions for the uncertainty H_n are considered below:

$$H_n := Var_n(\alpha)$$
$$\widetilde{H}_n := \int_D p_n(1 - p_n)\mathrm{d}\mu$$

Uncertainties:

$$H_n := Var_n(\alpha)$$
$$\widetilde{H}_n := \int_{\mathbb{X}} p_n(1 - p_n)\mathrm{d}\mu$$

SUR criteria:

$$J_n(\mathbf{x}) := \mathbb{E}_n(Var_{n+1}(\alpha))$$
$$\widetilde{J}_n(\mathbf{x}) := \mathbb{E}_n\left(\int_D p_{n+1}(1 - p_{n+1})\mathrm{d}\mu\right)$$

Main challenge to calculate $\widetilde{J}_n(\mathbf{x})$ (similar for $J_n(\mathbf{x})$): Obtain a closed form expression for $\mathbb{E}_n\left(p_{n+1}(1 - p_{n+1})\right)$ and integrate it.

Deriving SUR criteria

Proposition

$$\mathbb{E}_n(p_{n+1}(\mathbf{x})(1-p_{n+1}(\mathbf{x}))) = \Phi_2\left(\begin{pmatrix} a(\mathbf{x}) \\ -a(\mathbf{x}) \end{pmatrix}, \begin{pmatrix} c(\mathbf{x}) & 1-c(\mathbf{x}) \\ 1-c(\mathbf{x}) & c(\mathbf{x}) \end{pmatrix}\right)$$

- $\Phi_2(\cdot, M)$*: c.d.f. of centred bivariate Gaussian with covariance matrix* M
- $a(\mathbf{x}) := (m_n(\mathbf{x}) - t)/s_{n+1}(\mathbf{x})$,
- $c(\mathbf{x}) := s_n^2(\mathbf{x})/s_{n+1}^2(\mathbf{x})$

C. Chevalier, J. Bect, D. Ginsbourger, V. Picheny, E. Vazquez and Y. Richet.
Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set.
Technometrics, 56(4):455-465, 2014.

Deriving SUR criteria

Proposition

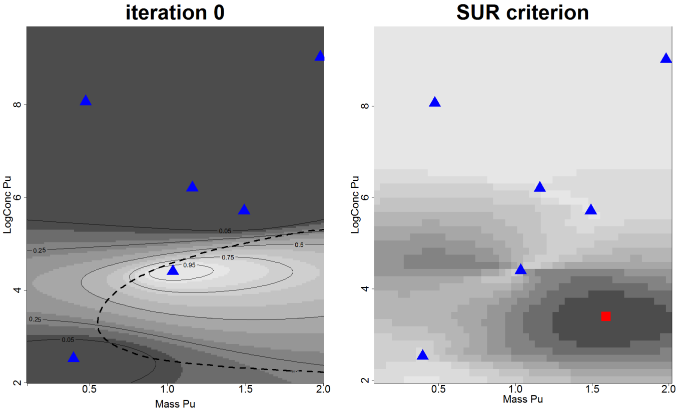
$$\mathbb{E}_n(p_{n+1}(\mathbf{x})(1 - p_{n+1}(\mathbf{x}))) = \Phi_2\left(\begin{pmatrix} a(\mathbf{x}) \\ -a(\mathbf{x}) \end{pmatrix}, \begin{pmatrix} c(\mathbf{x}) & 1 - c(\mathbf{x}) \\ 1 - c(\mathbf{x}) & c(\mathbf{x}) \end{pmatrix}\right)$$

- $\Phi_2(\cdot, M)$*: c.d.f. of centred bivariate Gaussian with covariance matrix* M
- $a(\mathbf{x}) := (m_n(\mathbf{x}) - t)/s_{n+1}(\mathbf{x})$,
- $c(\mathbf{x}) := s_n^2(\mathbf{x})/s_{n+1}^2(\mathbf{x})$

C. Chevalier, J. Bect, D. Ginsbourger, V. Picheny, E. Vazquez and Y. Richet.
Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set.
Technometrics, 56(4):455-465, 2014.

C. Chevalier, V. Picheny and D. Ginsbourger.
The KrigInv package: An efficient and user-friendly R implementation of Kriging-based inversion algorithms.
Computational Statistics & Data Analysis, 71:1021-1034, 2014.

Back to the test case with SUR



Batch-sequential SUR strategies

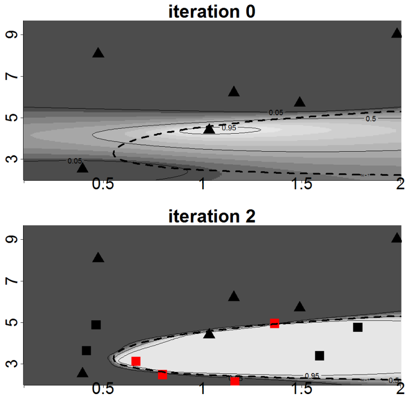


Figure: 3 SUR iterations ($\widetilde{J}_n$ criterion with $q = 4$)

Further questions about SUR and UQ on sets

About the consistency:

J. Bect, F. Bachoc and D. Ginsbourger (2018+).
A supermartingale approach to Gaussian process based sequential design of experiments.
HAL/Arxiv paper (hal-01351088, Arxiv: 1608.01118).

Further questions about SUR and UQ on sets

About the consistency:

J. Bect, F. Bachoc and D. Ginsbourger (2018+).
A supermartingale approach to Gaussian process based sequential design of experiments.
HAL/Arxiv paper (hal-01351088, Arxiv: 1608.01118).

Of course, in operational conditions, asymptotic results are worthwhile. However, concrete finite-sample outputs such as estimates of $\Gamma^\star$ and quantifications of the associated uncertainty are required as well.

Further questions about SUR and UQ on sets

About the consistency:

J. Bect, F. Bachoc and D. Ginsbourger (2018+).
A supermartingale approach to Gaussian process based sequential design of experiments.
HAL/Arxiv paper (hal-01351088, Arxiv: 1608.01118).

Of course, in operational conditions, asymptotic results are worthwhile. However, concrete finite-sample outputs such as estimates of $\Gamma^{\star}$ and quantifications of the associated uncertainty are required as well.

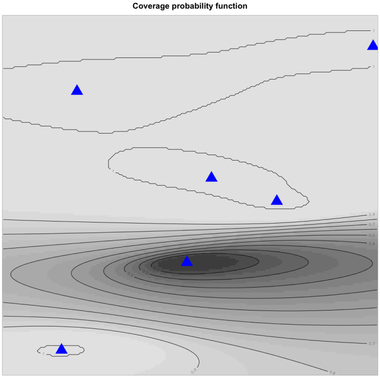
Now, n being fixed, how to estimate $\Gamma^{\star}$ and to assess/represent the variability of the corresponding estimate(s)?

How to summarize the posterior distribution of sets?

For application purposes, let us reverse the perspective and focus on the sojourn/excursion case **below** t, where $\Gamma = \{\mathbf{x} \in D : Z(\mathbf{x}) \leq t\}$ and $p_n : \mathbf{x} \in D \to p_n(x) = P_n(Z(\mathbf{x}) \leq t)$.

How to summarize the posterior distribution of sets?

For application purposes, let us reverse the perspective and focus on the sojourn/excursion case **below** t, where $\Gamma = \{\mathbf{x} \in D : Z(\mathbf{x}) \leq t\}$ and $p_n : \mathbf{x} \in D \to p_n(x) = P_n(Z(\mathbf{x}) \leq t)$.



Define the (conditional) quantiles of Γ as $\rho-$level sets of p_n:

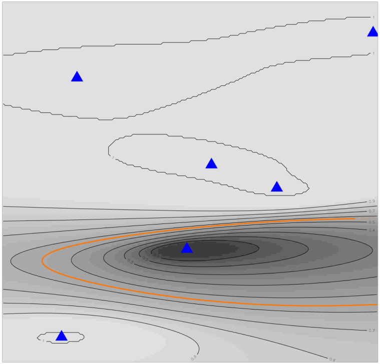
$$\begin{aligned} Q_\rho :&= \{\mathbf{x} \in D : p_n(\mathbf{x}) \geq \rho\} \\ &= \{\mathbf{x} \in D : P_n(Z(\mathbf{x}) \leq t) \geq \rho\}. \end{aligned}$$

How well Q_ρ estimates Γ can be quantified for instance through the “expected deviation”:

$$\mathbb{E}_n\left(\mu(Q_\rho \Delta \Gamma)\right)$$

Estimates of $\Gamma^{\star}$: the Vorob'ev expectation

Coverage probability function (Vorob'ev expectation)



The **Vorob'ev expectation of** $\Gamma \mid (Z_{x_1} = f(x_1), \dots, Z_{x_n} = f(x_n))$ is the $\rho^{\star}$ level set of p_n such that

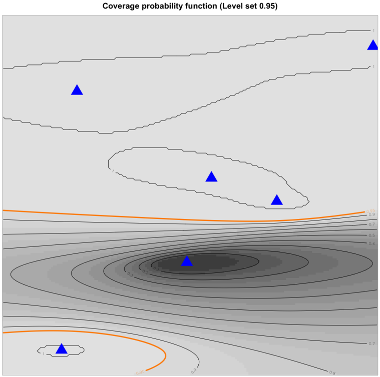
$$\mu(Q_{\rho^{\star}}) = \mathbb{E}_n[\mu(\Gamma)].$$

It is a state of the art result that $Q_{\rho^{\star}}$ minimizes $S \to \mathbb{E}_n(\mu(S\Delta\Gamma))$ among all closed sets $S \subset \mathbb{R}^d$ with volume $\mathbb{E}_n[\mu(\Gamma)]$.

C. Chevalier, D. Ginsbourger, J. Bect, and Molchanov, I.
Estimating and quantifying uncertainties on level sets using the Vorob'ev expectation and deviation with Gaussian process models.
mODa 10 Advances in Model-Oriented Design and Analysis, Physica-Verlag HD, 2013.

Estimates of $\Gamma^{\star}$: some limitations of Q_ρ quantiles

In practice one often wish to give **confidence statements** on the estimates.



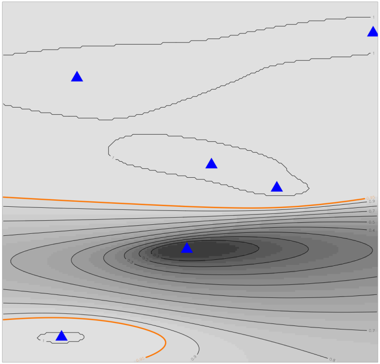
Q_ρ contains points which have marginal probability at least ρ of being in Γ.

$\Rightarrow$ no confidence statement on the probability of the actual excursion set containing this specific estimate.

Estimates of $\Gamma^\star$: some limitations of Q_ρ quantiles

In practice one often wish to give **confidence statements** on the estimates.

Coverage probability function (Level set 0.95)



Q_ρ contains points which have marginal probability at least ρ of being in Γ.

$\Rightarrow$ no confidence statement on the probability of the actual excursion set containing this specific estimate.

E.g., the probabilities of Q_ρ containing the excursion set (computed on a grid) are

- 0.67 for $\rho = 0.95$
- 0.009 for $\rho = 0.5$
- 0.019 for $\rho = 0.56$ (Vorob'ev)

Conservative Estimates of $\Gamma^\star$

We denote by **conservative estimate** for $\Gamma \mid (Z_{x_1} = f(x_1), \ldots, Z_{x_n} = f(x_n))$ at level β the largest Q_ρ such that $\mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta$:

$$E_{t,\alpha} = \arg\max_{Q_\rho} \{\mu(Q_\rho) : \mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta\}$$

D. Bolin, F. Lindgren.
Excursion and contour uncertainty regions for latent Gaussian models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014.

Conservative Estimates of $\Gamma^\star$

We denote by **conservative estimate** for $\Gamma \mid (Z_{x_1} = f(x_1), \ldots, Z_{x_n} = f(x_n))$ at level β the largest Q_ρ such that $\mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta$:

$$E_{t,\alpha} = \arg\max_{Q_\rho} \{\mu(Q_\rho) : \mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta\}$$

D. Bolin, F. Lindgren.
Excursion and contour uncertainty regions for latent Gaussian models.
Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2014.

Such conservative estimate $E_{t,\beta}$ is hence

- the largest quantile such that, with probability β, the response is below the threshold simultaneously at each of its locations.
- based on a confidence statement on the whole set

Computing conservative estimates

The computation of a conservative estimate

$$E_{t,\beta} = \arg\max_{Q_\rho}\{\mu(Q_\rho) : \mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta\}$$

presents two (nested) computational bottlenecks:

1. find the set with the maximum volume;
2. compute $\mathbb{P}_n(Q_\rho \subset \Gamma)$.

Computing conservative estimates

The computation of a conservative estimate

$$E_{t,\beta} = \arg\max_{Q_\rho} \{\mu(Q_\rho) : \mathbb{P}_n(Q_\rho \subset \Gamma) \geq \beta\}$$

presents two (nested) computational bottlenecks:

1. find the set with the maximum volume;
2. compute $\mathbb{P}_n(Q_\rho \subset \Gamma)$.

For recent work on computing the last term, see for instance

D. Azzimonti and D. Ginsbourger (2018).
Estimating orthant probabilities of high dimensional Gaussian vectors with an application to set estimation.
Journal of Computational and Graphical Statistics, 27:2, 255-267

Computing $\mathbb{P}_n(Q_\rho \subset \Gamma)$

If Q_ρ is discretized over a grid $W = \{w_1, \ldots, w_m\}$, then

$$\mathbb{P}_n(Q_\rho \subset \Gamma) = \mathbb{P}_n(Z_{w_1} \leq t, \ldots, Z_{w_m} \leq t) = 1 - \mathbb{P}_n\left(\max_{i=1,\ldots,m} Z_{w_i} > t\right)$$

Computing $\mathbb{P}_n(Q_\rho \subset \Gamma)$

If Q_ρ is discretized over a grid $W = \{w_1, \dots, w_m\}$, then

$$\mathbb{P}_n(Q_\rho \subset \Gamma) = \mathbb{P}_n(Z_{w_1} \leq t, \dots, Z_{w_m} \leq t) = 1 - \mathbb{P}_n\left(\max_{i=1,\dots,m} Z_{w_i} > t\right)$$

There exists a number of algorithms to estimate $\mathbb{P}_n(Z_{w_1} \leq t, \dots, Z_{w_m} \leq t)$:

1. quasi-MC integration techniques
   - very fast and reliable in small dimensions;
   - hardly usable for dimensions higher than 1000.
2. pure MC techniques:
   - dimension independent;
   - high number of simulations for small variance.

Computing $\mathbb{P}_n(Q_\rho \subset \Gamma)$

If Q_ρ is discretized over a grid $W = \{w_1, \ldots, w_m\}$, then

$$\mathbb{P}_n(Q_\rho \subset \Gamma) = \mathbb{P}_n(Z_{w_1} \leq t, \ldots, Z_{w_m} \leq t) = 1 - \mathbb{P}_n\left(\max_{i=1,\ldots,m} Z_{w_i} > t\right)$$

There exists a number of algorithms to estimate $\mathbb{P}_n(Z_{w_1} \leq t, \ldots, Z_{w_m} \leq t)$:

1. quasi-MC integration techniques
   - very fast and reliable in small dimensions;
   - hardly usable for dimensions higher than 1000.
2. pure MC techniques:
   - dimension independent;
   - high number of simulations for small variance.

IRSN test case

- an estimate with a good resolution requires an 100×100 grid for D;
- W consists of $+1000$ grid points for some Q_ρ.

$\mathbb{P}_n(\max_{w \in W} Z_w > T)$: proposed hybrid algorithm

Algorithm:

1. select q grid points, denoted $W_q \subset W$;
2. compute $p' = P(\max_{w \in W_q} Z_w > t)$ with qMC quadrature;
3. estimate $\mathbb{P}_n(\max_{w \in W} Z_w > t)$ with

$$\hat{p} = p' + (1 - p')\hat{R}_q$$

where $\hat{R}_q$ is a MC estimator of

$$R_q = \mathbb{P}_n\left(\max_{w \in W \setminus W_q} Z_w > t \,\middle|\, \max_{w \in W_q} Z_w \leq t\right)$$

$\mathbb{P}_n(\max_{w \in W} Z_w > T)$: proposed hybrid algorithm

Algorithm:

1. select q grid points, denoted $W_q \subset W$;
2. compute $p' = P(\max_{w \in W_q} Z_w > t)$ with qMC quadrature;
3. estimate $\mathbb{P}_n(\max_{w \in W} Z_w > t)$ with

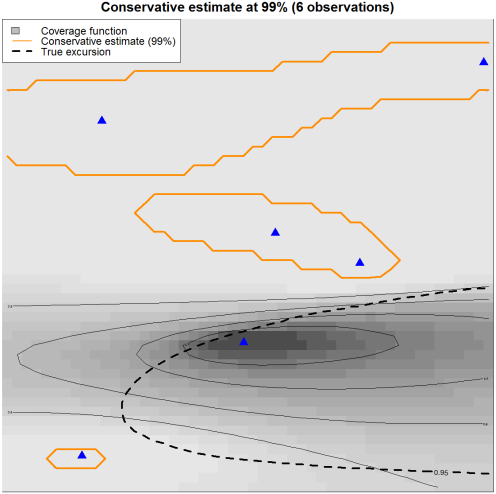
$$\hat{p} = p' + (1 - p')\hat{R}_q$$

where $\hat{R}_q$ is a MC estimator of

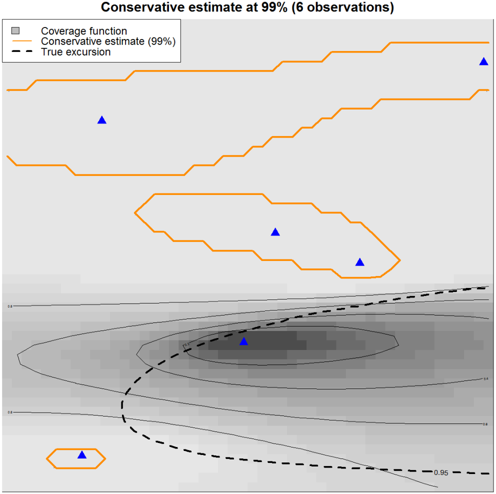
$$R_q = \mathbb{P}_n\left(\max_{w \in W \setminus W_q} Z_w > t \,\middle|\, \max_{w \in W_q} Z_w \leq t \right)$$

An asymmetric nested Monte Carlo scheme was developed for improved efficiency in R_q's estimation. (See "orthant" paper and **anMC** R package).

Back to the test case with a conservative estimate. . .

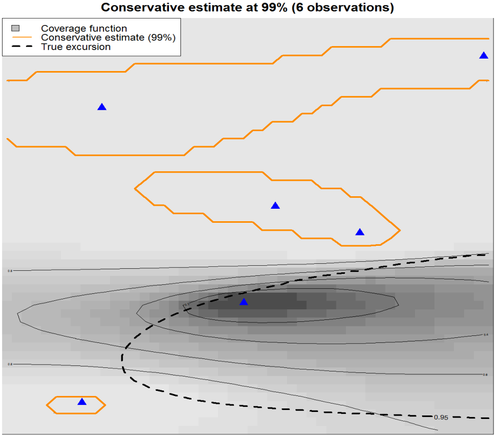


Back to the test case with a conservative estimate. . .



NB: here, $\rho = 99.88829\%$ for a confidence of 99.12178%.

. . . and associated sequential strategies



For more on sequential conservative estimation

D. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, Y. Richet (2018+).
Adaptive Design of Experiments for Conservative Estimation of Excursion Sets.
arXiv:1611.07256v2 [stat.ME]

For more on sequential conservative estimation

D. Azzimonti, D. Ginsbourger, C. Chevalier, J. Bect, Y. Richet (2018+).
Adaptive Design of Experiments for Conservative Estimation of Excursion Sets.
arXiv:1611.07256v2 [stat.ME]

Some open questions and perspectives

- Asymptotic results in the conservative case?
- Study the effect of threshold plug-in in the criteria.
- Investigating options closer to ”Full Bayesian” for this problem.

Overall perspectives on GP-based set estimation

- Transpose work to other families of implicitly defined regions.
- Consider families of set estimates beyond quantiles.
- Investigate rates of convergence for SUR strategies (?).

Overall perspectives on GP-based set estimation

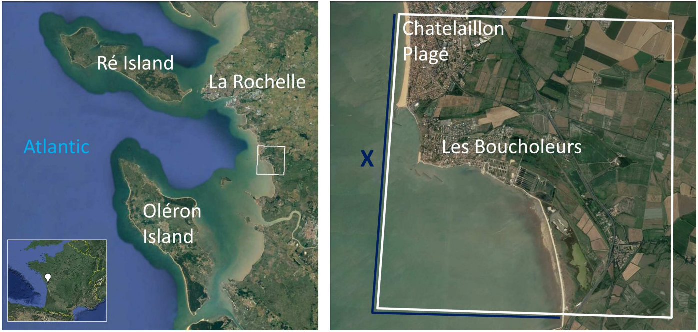
- Transpose work to other families of implicitly defined regions.
- Consider families of set estimates beyond quantiles.
- Investigate rates of convergence for SUR strategies (?).

Acknowledgements:

Drs Yann Richet and Grégory Caplin (French Nuclear Safety Institute) for providing the criticality safety test case.

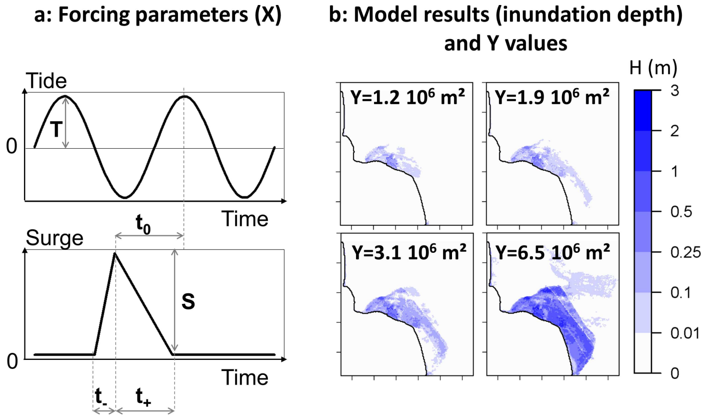
Special thanks to Drs. Dario Azzimonti and Clément Chevalier for numerous invaluable inputs, and more generally, to all co-authors involved.

Simulation of coastal flooding at “Les Boucholeurs”



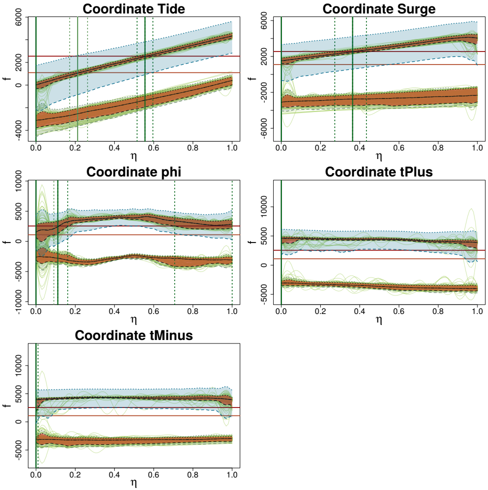
Study site location (left) and computational domain limits (right, in white) with location of the forcing conditions (right, in blue).

Test case input and output parametrization



(a) Schematic representation of the tide and surge temporal signals and the different parameters describing them. (b) Maps of inland water height for given values of the parameters, and deduced value of flood surface.

Estimated coordinate profile maxima for the 5 inputs



Key underlying result

Theorem

Consider $(Z_{\mathbf{x}})_{\mathbf{x}\in D} \sim GP(\mu, \mathfrak{K})$ and an approximating process of Z, $\widetilde{Z}$, defined by $\widetilde{Z}_{\mathbf{x}} = a(\mathbf{x}) + \mathbf{b}^T(\mathbf{x})Z_G$ where the $\mathbf{a}, \mathbf{b}$ functions and $G = \{\mathbf{g}_1, \ldots, \mathbf{g}_\ell\} \subset D$ ($\ell \geq 1$) are given. Then, for $T \subset D$ and any $u > \mu_T^{\widetilde{\Delta}}$,

$$\mathbb{P}\left(|\sup_{\mathbf{x}\in T} Z_{\mathbf{x}} - \sup_{\mathbf{x}\in T} \widetilde{Z}_{\mathbf{x}}| > u\right) \leq 2\exp\left(-\frac{(u-\mu_T^{\widetilde{\Delta}})^2}{2(\sigma_T^{\widetilde{\Delta}})^2}\right), \tag{1}$$

where

$$\mu_T^{\widetilde{\Delta}} = \sup_{\mathbf{x}\in T}|\mu^{\widetilde{\Delta}}(\mathbf{x})| \text{ and } (\sigma_T^{\widetilde{\Delta}})^2 = \sup_{\mathbf{x}\in T} \mathfrak{K}^{\widetilde{\Delta}}(\mathbf{x},\mathbf{x}) \text{ with} \tag{2}$$

$$\mu^{\widetilde{\Delta}}(\mathbf{x}) = \mathbb{E}[Z_{\mathbf{x}} - \widetilde{Z}_{\mathbf{x}}] = \mu(\mathbf{x}) - a(\mathbf{x}) - \mathbf{b}^T(\mathbf{x})\mu(G)$$

$$\mathfrak{K}^{\widetilde{\Delta}}(\mathbf{x},\mathbf{x}') = \mathfrak{K}(\mathbf{x},\mathbf{x}') - \mathfrak{K}(\mathbf{x}',G)\mathbf{b}(\mathbf{x}) - \mathfrak{K}(\mathbf{x},G)\mathbf{b}(\mathbf{x}') + \mathbf{b}^T(\mathbf{x})\mathfrak{K}(G,G)\mathbf{b}(\mathbf{x}'),$$

If $\widetilde{Z} - Z$ is centred then (1) is valid for any $u > 0$.

For more detail

More on the profile maxima approach and its application to the BRGM data can be found in

D. Azzimonti, D. Ginsbourger, J. Rohmer, D. Idier (2017+)
Profile extrema for visualizing and quantifying uncertainties on excursion regions. Application to coastal flooding.
https://arxiv.org/abs/1710.00688

For more detail

More on the profile maxima approach and its application to the BRGM data can be found in

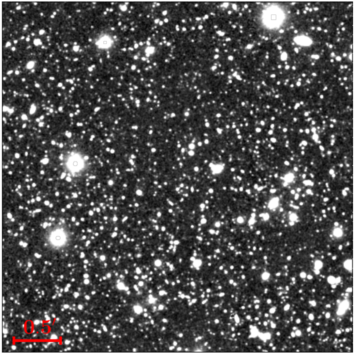
D. Azzimonti, D. Ginsbourger, J. Rohmer, D. Idier (2017+)
Profile extrema for visualizing and quantifying uncertainties on excursion regions. Application to coastal flooding.
https://arxiv.org/abs/1710.00688

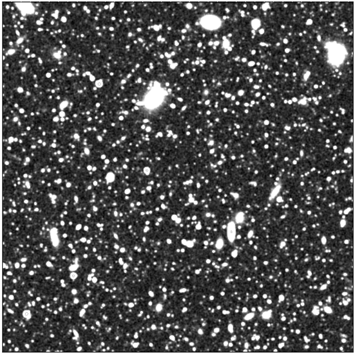
For more on random fields and geometry, see in particular

R. J. Adler and J. E. Taylor (2007)
Random Fields and Geometry.
Springer

and references therein.

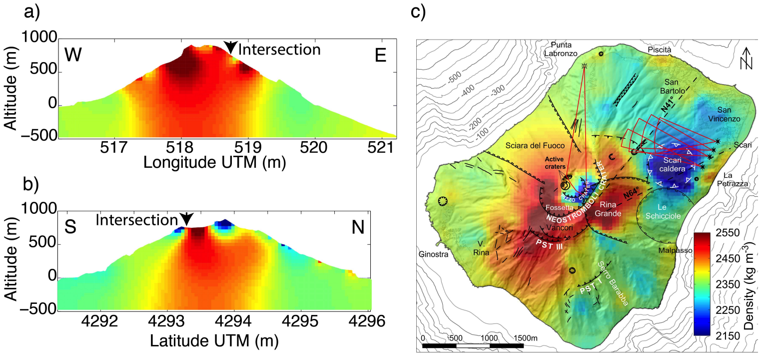
Reducing uncertainties on cosmological parameters





Comparison of observed astronomical data (left) versus a simulated image (right) corresponding to one of the posterior samples in the approximate Bayesian inversion approach employed in Herbel et al. 2017.

Sequential design to locate past volcano activity



(a-b) Vertical sections of the inferred 3-D density of Stromboli. (c) Aerial view of the shallow density distribution with superimposed topography and geological interpretation. Modified from Linde et al 2014.

Generalized optimality property for Vorob'ev quantiles

Proposition

For any $\rho \in [0, 1]$, the Vorob'ev quantile

$$Q_\rho = \{x \in D : p_n(x) \geq \rho\}$$

minimizes the expected distance in measure with Γ among measurable sets M such that $\mu(M) = \mu(Q_\rho)$, i.e.,

$$\mathbb{E}_n\left[\mu(Q_\rho \Delta \Gamma)\right] \leq \mathbb{E}_n\left[\mu(M \Delta \Gamma)\right],$$

for any measurable set M such that $\mu(M) = \mu(Q_\rho)$.

Generalized optimality property for Vorob'ev quantiles

Proposition

For any $\rho \in [0, 1]$, the Vorob'ev quantile

$$Q_\rho = \{x \in D : p_n(x) \geq \rho\}$$

minimizes the expected distance in measure with Γ among measurable sets M such that $\mu(M) = \mu(Q_\rho)$, i.e.,

$$\mathbb{E}_n \left[\mu(Q_\rho \Delta \Gamma)\right] \leq \mathbb{E}_n \left[\mu(M \Delta \Gamma)\right],$$

for any measurable set M such that $\mu(M) = \mu(Q_\rho)$.

A proof of this property is presented in Dario Azzimonti's PhD thesis (2016).