Optimal design for targeted region in Gaussian Fields

Pierre Druilhet

(joint work with Sylvain Coly)

Laboratoire de Mathématiques Blaise Pascal, UMR CNRS 6620
Université Clermont Auvergne

Framework

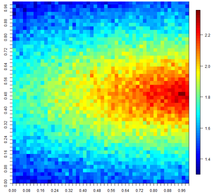
Framework

- Design of Experiments.
- Monitoring framework (environment, health, climate, . . .). Examples : Pollution of a lake, volcanic eruption, etc.
- Approximately known phenomenon.
- Acquisition of spatial data : sensor positions (fixed or mobile).

$\longrightarrow$ More accurate knowledge in "critical" areas.

Meta-model

Mapping of $y(x)$



- E is an $n \times n$ grid
- $y(x)$ value of the variable of interest for $x \in E$;
- $y \sim \mathcal{N}(\mu, \Sigma)$ is a Gaussian field ;
- Σ is a known covariance matrix .

k-point design

Aim

Getting information on $y(x)$ by observing k points : $y(x_1), ..., y(x_n)$

How to choose the points ?

- First, specify the goal of the experiment.
- Then define a criterion.
- At last construct the k-point design $\mathcal{D}^* = \{x_1^*, ..., x_k^*\}$ that minimizes the criterion.

Updating formula

From a k-point design $\mathcal{D} = \{x_1, ..., x_k\}$, we get the observations $y_{\mathcal{D}} = \{y_{x_1}, ..., y_{x_k}\}$.
For $x \notin \mathcal{D}$, we get an updated mean and variance :

Updated mean

$$\mu(x) \mid y_{\mathcal{D}} = \mu(x) + \mathrm{Cov}\left((y(x); y_{\mathcal{D}}\right) \cdot \left\{\mathrm{Var}(y_{\mathcal{D}})\right\}^{-1} \cdot (y_{\mathcal{D}} - \mu_{\mathcal{D}})$$

Updated Variance

$$\mathrm{Var}\left(y(x) \,|\, y_{\mathcal{D}}\right) = \mathrm{Var}(y(x)) - \mathrm{Cov}\left(y(x); y_{\mathcal{D}}\right) \cdot \left\{\mathrm{Var}(y_{\mathcal{D}})\right\}^{-1} \cdot \mathrm{Cov}\left(y_{\mathcal{D}}; y(x)\right)$$

The updated variance does not depend on the observations $y_{\mathcal{D}}$, but only on the location of the design points $\mathcal{D}$!

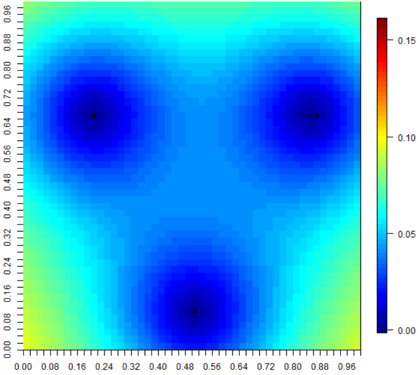
Example of updated variance

3-point design.

The variance decreases all around these points.

When a new point is added, the covariance matrix can be updated.

Map of variance.

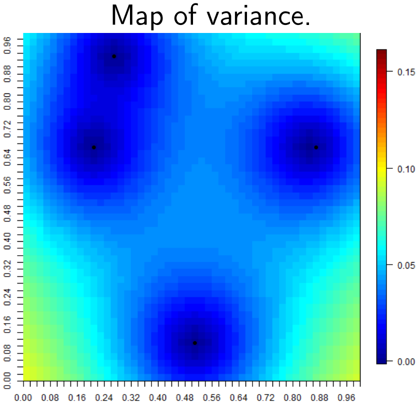


Example of updated variance

4-point design.

The variance decreases all around these points.

When a new point is added, the covariance matrix can be updated.



Space-filling designs

Aim

Control the variance over the whole space.

Related possible criteria

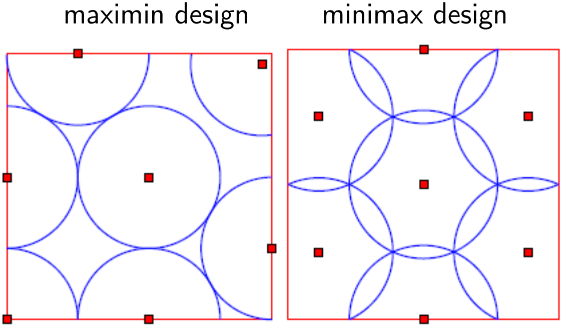
- $MC(\mathcal{D}) = \max_x \mathrm{Var}(y(x)|\mathcal{D})$ (Max criterion)
- $IC(\mathcal{D}) = \sum_{x \in E} \mathrm{Var}(y(x)|\mathcal{D})$ (Integrated criterion)

Optimization algorithm are time consuming !

Alternative criteria for space-filling designs

- $C(\mathcal{D}) = \max_{x \in E} \; \min_{i=1,\ldots,k} d(x, x_i)$ $\hookrightarrow$ minimax design
- $C(\mathcal{D}) = -\min_{i,j=1,..,k} d(x_i, x_j)$ $\hookrightarrow$ maximin design

Example of 7-point space-filling design



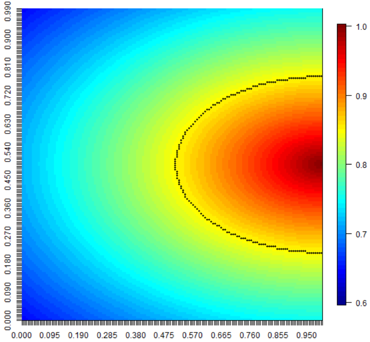
Other space filling designs

- **Hypercube latin designs** [Stein (1987)], **orthogonal arrays** [Owen (1992)].
- **Low discrepancy sequences** (Halton, Hammersley, Sobol, Faure).
- **Optimal designs (estimation) :** Maximum entropy [Currin et *al.* (1991)], IMSE [Sacks et *al.* (1989)].
- **Packages R :** DiceDesign [Franco, Dupuy et *al.* (2015)], randtoolbox [Chalabi et *al.* (2014)].

Design for targeted area

Aim 1 : Exploring area with high values for $y(x)$

Map of $y(x)$.



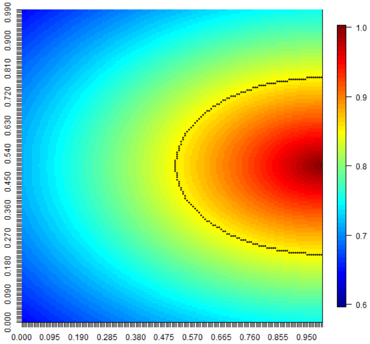
We are interested in controlling the variance over the region exceeding a given threshold T

$$R_T^{exc} = \{x \;/\; y(x) > T\}\,.$$

Design for targeted area

Aim 2 : Level line detection

Map of $y(x)$.



We are interested in detecting the level line for a given threshold T

$$R_T^{LL} = \{x \;/\; y(x) = T\}\,.$$

Sequential vs non-sequential design

- For a non-sequential design, we choose the k points before the experiment. We must rely on $\mu(x)$ to focus on the area of interest !
- For a sequential design, after each point (or a group of points), we perform measurements. We choose the next point(s) as for the non-sequential design but based on the updated $\mu(x) = \mathbb{E}(y(x)|y_{\mathcal{D}})$ and $\mathrm{var}(y(x)|\mathcal{D})$.

Criterion for targeted area

Weighted variance

To target a given area , the general principal is to a put more weight on the point that are possibly in the zone of interest and to control the variance.

$$c(x;\mathcal{D}) = \mathrm{weight}(x) \times \mathrm{var}(y(x)|\mathcal{D})$$

Derived criteria for a design $\mathcal{D}$

- Max criterion

$$MC(\mathcal{D}) = \max_{x} \mathrm{weight}(x) \times \mathrm{var}(y(x)|\mathcal{D})$$

- Integrated criterion

$$IC(\mathcal{D}) = \sum_{x} \mathrm{weight}(x) \times \mathrm{var}(y(x)|\mathcal{D})$$

Example 1

Non-sequential design exploring area with high values for $y(x)$

Target region

$$R_T^{exc} = \{x \;/\; y(x) > T\}$$

Weight

$$p_T^{\mathrm{exc}}(x) = F\left(\frac{\mu_x - T}{\sigma_x}\right) \quad (1)$$

where F is the cumulative density function of a $\mathcal{N}(0,1)$

Derived criteria

- Integrated criterion : $IC_T^{\mathrm{exc}}(\mathcal{D}) = \sum_x p_T^{\mathrm{exc}}(x).\mathrm{Var}\left(y(x)|\mathcal{D}\right).$
- max criterion : $MC_T^{\mathrm{exc}}(\mathcal{D}) = \max_x p_T^{\mathrm{exc}}(x).\mathrm{Var}\left(y(x)|\mathcal{D}\right).$

Interpretation of the weight $p_T^{\text{exc}}(x)$ as a probability

$$p_T^{\text{exc}}(x;\mathcal{D}) = \mathbb{P}(y(x) > T)$$

Interpretation of the weight $p_T^{\text{exc}}(x)$ as a p-value

Consider :

- $y(x)$ as an unknown fixed quantity ;
- $\mu_x \sim \mathcal{N}(y(x), \mathrm{Var}(y(x)))$.

Then, $p_T^{\text{exc}}(x)$ appears as the p-value of the test

$$\mathcal{H}_0 : \text{“}y(x) > T\text{”} \text{ vs } \mathcal{H}_1 : \text{“}y(x) \leq T\text{”}$$

General method for constructing an optimal computer design

1. Constructing an initial design $\mathcal{D}_0$
2. Improving this design by an exchange algorithm [Kennard et Stone (1969), Fedorov (1972)].

Step 1 : Construction of the initial $\mathcal{D}^{(0)} = \left(x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}\right)$

- $x_1^{(0)} = \mathrm{ArgMax}_{x \in E} \left(c_T^{\mathrm{thres}}(x; \mathcal{D}_0^{(0)} = \emptyset) \right) \longrightarrow \mathcal{D}_1^{(0)} = \left\{ x_1^{(0)} \right\};$
- $x_2^{(0)} = \mathrm{ArgMax}_{x \notin \mathcal{D}_1^{(0)}} \left(c_T^{\mathrm{thres}}(x; \mathcal{D}_1^{(0)}) \right) \longrightarrow \mathcal{D}_2^{(0)} = \mathcal{D}_1^{(0)} \cup \left\{ x_2^{(0)} \right\};$
- ....
- $x_i^{(0)} = \mathrm{ArgMax}_{x \notin \mathcal{D}_{i-1}^{(0)}} \left(c_T^{\mathrm{thres}}(x; \mathcal{D}_{i-1}^{(0)}) \right) \longrightarrow \mathcal{D}_i^{(0)} = \mathcal{D}_{i-1}^{(0)} \cup \left\{ x_i^{(0)} \right\};$
- ....
- $x_k^{(0)} = \mathrm{ArgMax}_{x \notin \mathcal{D}_{k-1}^{(0)}} \left(c_T^{\mathrm{thres}}(x; \mathcal{D}_{k-1}^{(0)}) \right) \longrightarrow \mathcal{D}^{(0)} = \mathcal{D}_{k-1}^{(0)} \cup \left\{ x_k^{(0)} \right\}.$

where $c_T^{\mathrm{thres}}(x; \mathcal{D}) = p_T^{\mathrm{exc}}(x) \times \mathrm{var}(y(x)|\mathcal{D})$ is the individual contribution of x to the criterion.

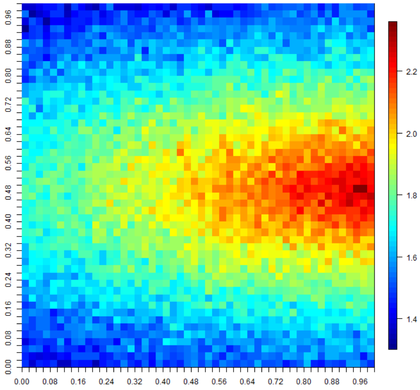
Step 2 : Apply the exchange algorithm.

Algorithm 1: Exchange algorithm

Input: initial design $\mathcal{D}^{(0)}$, maximal number of iterations M ;
foreach k *from* 1 *to* M **do**
    randomly draw $x \in \mathcal{D}^{(k-1)}$;
    randomly draw $x' \in E \backslash \mathcal{D}^{(k-1)}$;
    permute x and x' considering $\mathcal{D}^* = \mathcal{D}^{(k-1)} \cup \{x'\} \backslash \{x\}$;
    if $C_T(\mathcal{D}^*) < C_T(\mathcal{D}^{(k-1)})$ **then**
        $\mathcal{D}^{(k)} = \mathcal{D}^*$;
    else
        $\mathcal{D}^{(k)} = \mathcal{D}^{(k-1)}$;
    end
end
Output: plan $\mathcal{D}^*$.

Map of the $\mathcal{D}^{(0)}$ design

Map of $y(x)$.



Example

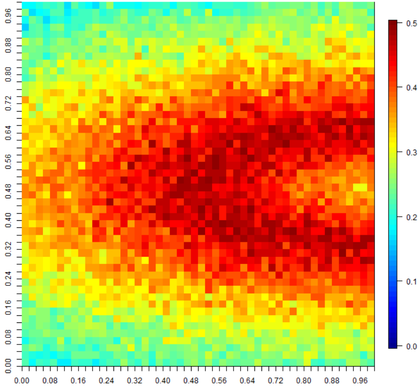
- Approximately known elliptical signal.
- 10-point designs.

Aim

To minimize the estimation error by targeting the level set defined by a threshold T.

Map of the $\mathcal{D}^{(0)}$ design

Initial map of $c_T^{\text{thres}}(x)$.



Example

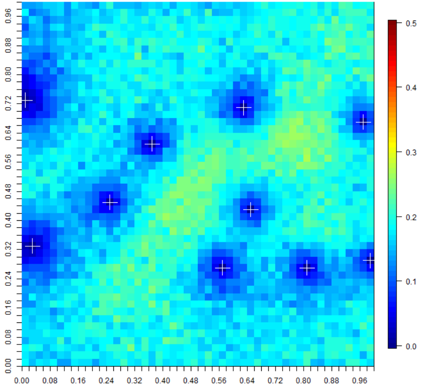
- Approximately known elliptical signal.
- 10-point designs.

Aim

To minimize the estimation error by targeting the level set defined by a threshold T.

Map of the $\mathcal{D}^{(0)}$ design

Map of $c_T^{\text{thres}}(c)$ for $\mathcal{D}^{(0)}$.



Example

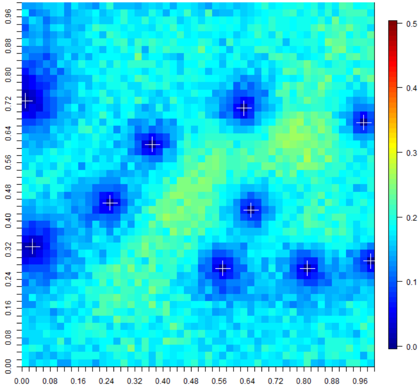
- Approximately known elliptical signal.
- 10-point designs.

Aim

To minimize the estimation error by targeting the level set defined by a threshold T.

MAP OF THE $\mathcal{D}^{(0)}$ DESIGN

Map of $c_T^{\text{thres}}(c)$ for $\mathcal{D}^{(10000)}$.



EXAMPLE

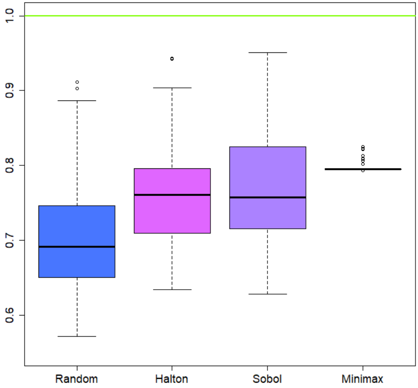
- Approximately known elliptical signal.
- 10-point designs.

AIM

To minimize the estimation error by targeting the level set defined by a threshold T.

- no accepted exchange (even with 10000 iter.).

Comparison between $\mathcal{D}^{(0)}$ and alternative designs



Initial designs

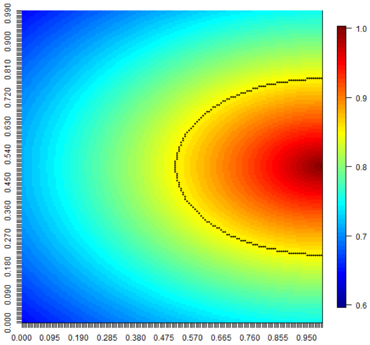
- $\mathcal{D}^{(0)}$ outperforms every other initial design.

After exchange algorithm

- Less than 1% of the resulting designs are better than $\mathcal{D}^{(0)}$.
- $\mathcal{D}^{(0)}$ remains very relevant ; its efficiency is about 0.998.

Example 2 : sequential design for level-line detection

Map of $y(x)$.



We are interested in detecting the level line

$$R_T^{LL} = \{x \; / \; y(x) = T\} .$$

Target region

$$R_T^{LL} = \{x \ / \ y(x) = T\}$$

Weight

$$p_T^{\text{thres}}(x) = 1 - 2\left|\frac{1}{2} - F\left(\frac{\mu_x - T}{\sigma_x}\right)\right|. \qquad (2)$$

where F is the cumulative density function of a $\mathcal{N}(0,1)$

Derived criteria

- Integrated criterion : $IC_T^{\text{thres}}(\mathcal{D}) = \sum_x p_T^{\text{thres}}(x).\mathrm{Var}\left(y(x)|\mathcal{D}\right).$
- Max criterion : $MC_T^{\text{thres}}(\mathcal{D}) = \max_x p_T^{\text{thres}}(x).\mathrm{Var}\left(y(x)|\mathcal{D}\right).$

Interpretation of the weight $p_T^{\text{thres}}(x)$ as a p-value

Consider :

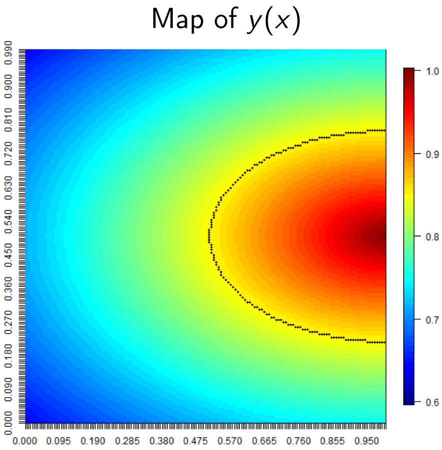
- $y(x)$ as an unknown fixed quantity ;
- $\mu_x \sim \mathcal{N}(y(x), \mathrm{Var}(y(x)))$.

Then, $p_T^{\text{thres}}(x)$ appears as the p-value of the bilateral test

$$\mathcal{H}_0 : \text{“}y(x) = T\text{”} \text{ vs } \mathcal{H}_1 : \text{“}y(x) \neq T\text{”}$$

Default choice for μ and starting design

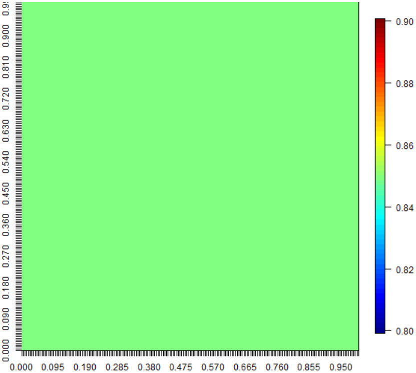
- In absence of prior information, we choose at first $\mu(x) = T$ so that each point have the maximum weight $= 1$
- We start with a space-filling design with few points to get information.
- Then we construct the next points by applying the algorithm described in the non-sequential design.



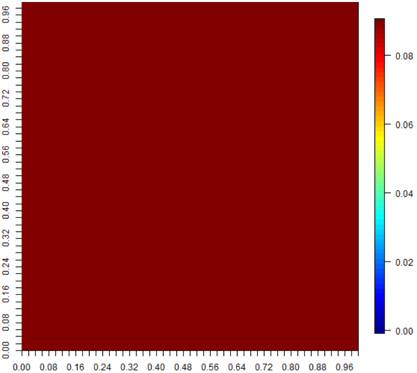
Goal : Find an optimal sequential design w.r.t. $MC_T^{\text{thres}}(\mathcal{D})$

Initial surrogate model and the value of c(x)

Map of $\mu(x) = T$

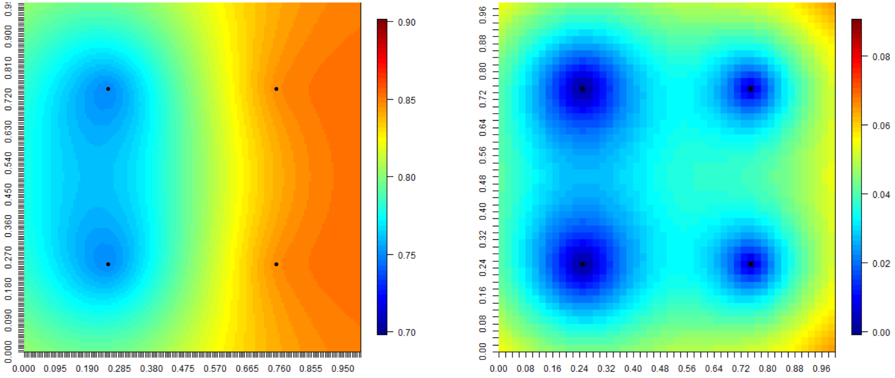


Map of $c^{\text{thres}}(x)$.



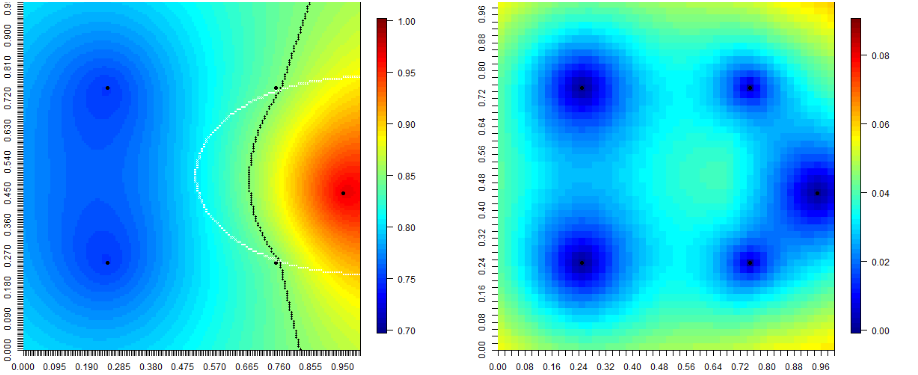
4-point initial space-filling design

Map of $\mu(x)$ Map of $c^{\text{thres}}(x)$.



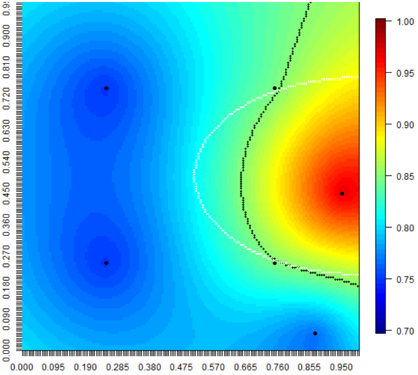
5-POINT DESIGN

Map of $\mu(x)$ | Map of $c^{\text{thres}}(x)$.

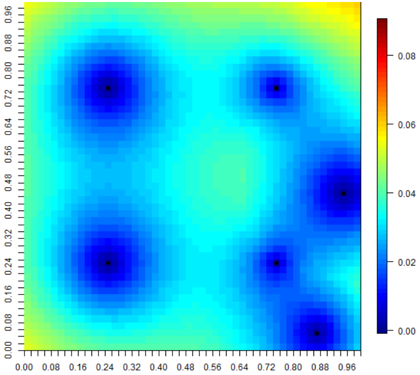


6-point design

Map of $\mu(x)$

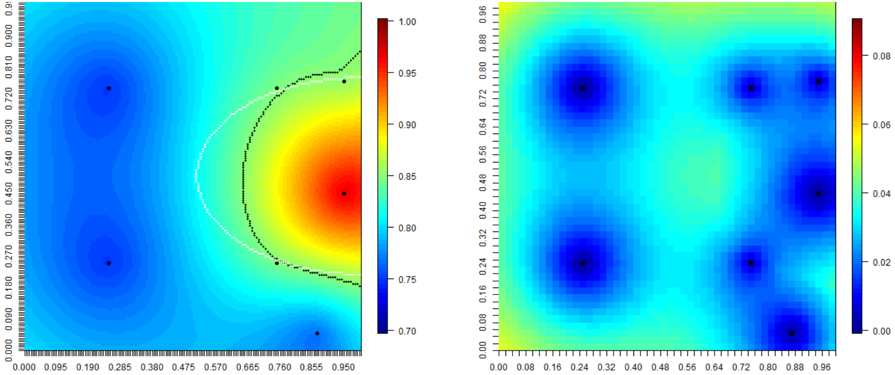


Map of $c^{\text{thres}}(x)$.



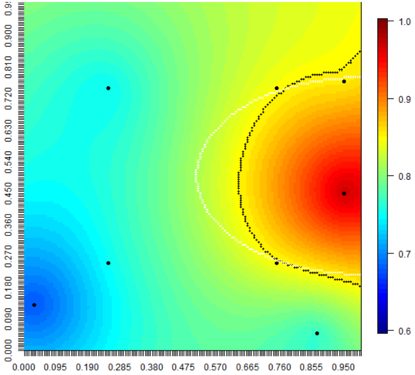
7-POINT DESIGN

Map of $\mu(x)$ Map of $c^{\text{thres}}(x)$.

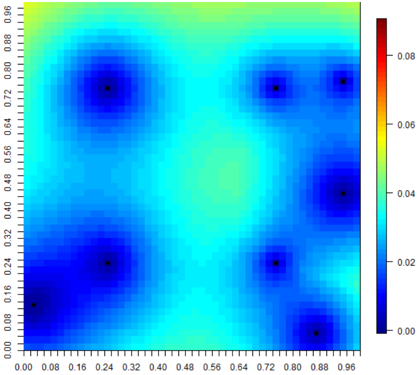


8-point design

Map of $\mu(x)$

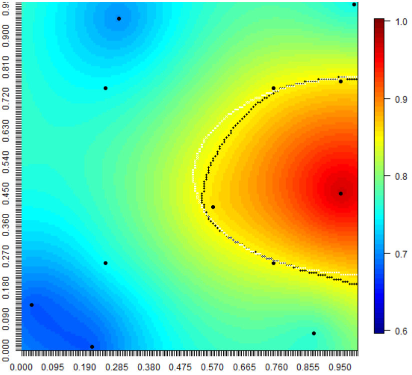


Map of $c^{\text{thres}}(x)$.

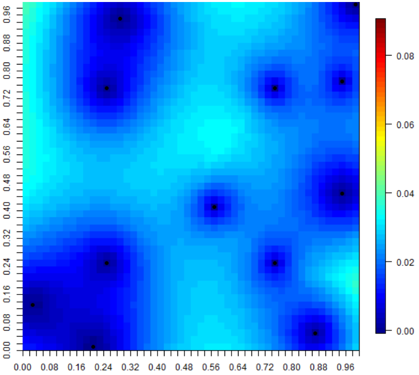


12-POINT DESIGN

Map of $\mu(x)$

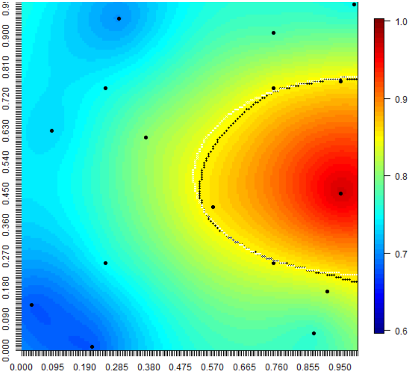


Map of $c^{\text{thres}}(x)$.

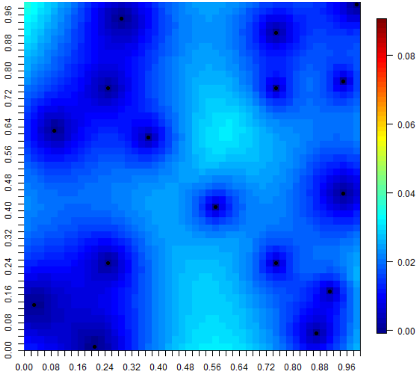


16-point design

Map of $\mu(x)$

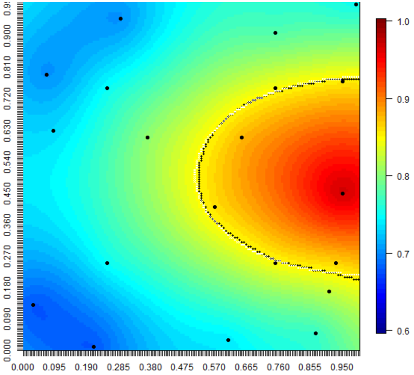


Map of $c^{\text{thres}}(x)$.

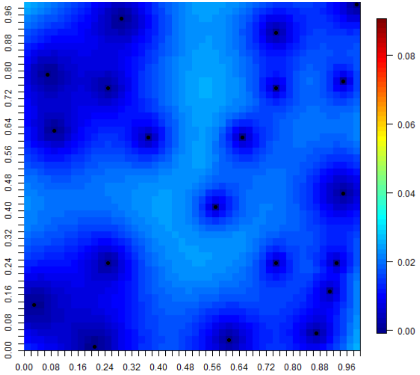


20-POINT DESIGN

Map of $\mu(x)$



Map of $c^{\text{thres}}(x)$.



Alternative weights and related criteria : target MSE

Picheny et *al.* (2010) proposed another weight given by

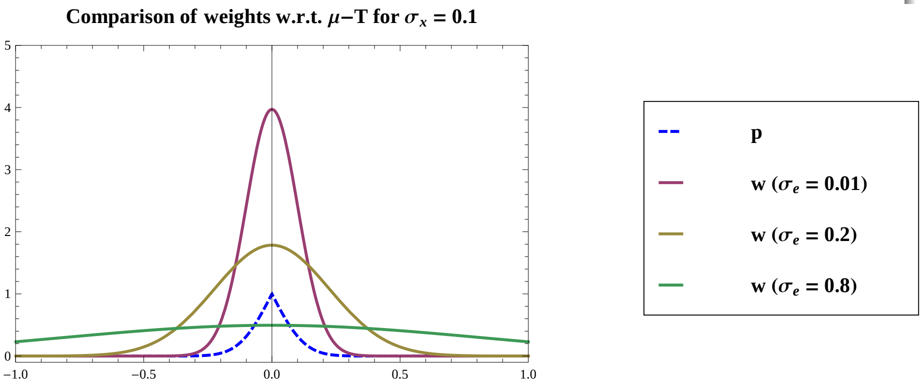
$$w_{\sigma_\epsilon^2}(x) = \frac{1}{\sqrt{2\pi\left(\sigma_\epsilon^2 + \sigma_x^2\right)}} \exp\left[-\frac{(\mu_x - T)^2}{2\left(\sigma_\epsilon^2 + \sigma_x^2\right)}\right].$$

where σ_ϵ need to be calibrated.
Note that $\sigma_\varepsilon = 0$ lead to unbounded weights when $\sigma_x \approx 0$.

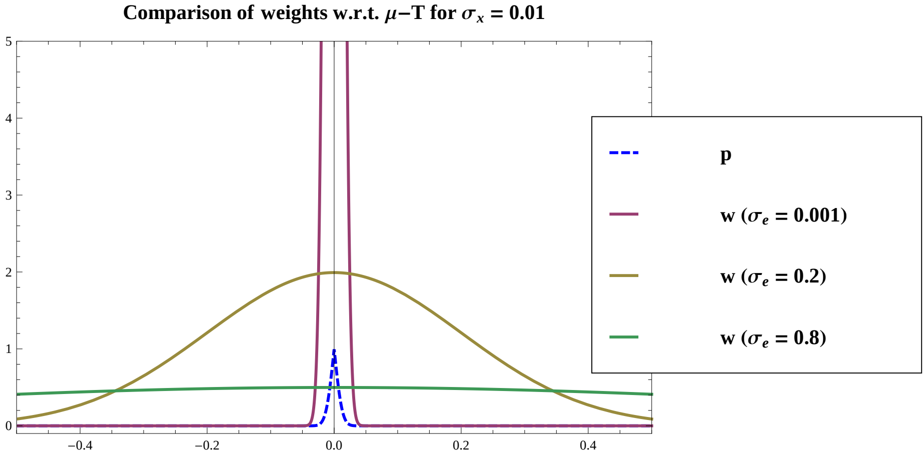
Related criteria

- tIMSE $= \sum_x w_{\sigma_\epsilon^2}(x)\,\mathrm{var}(y(x)|\mathcal{D})$.
- tMMSE $= \max_x w_{\sigma_\epsilon^2}(x)\,\mathrm{var}(y(x)|\mathcal{D})$

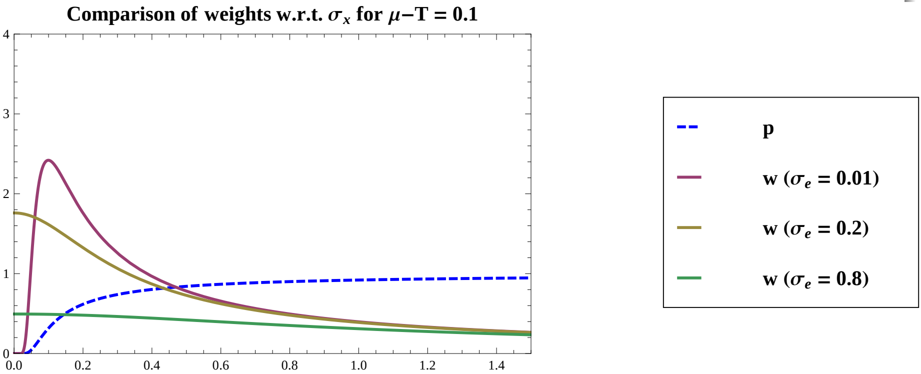
Comparison of weights w.r.t. $\mu(x) - T$ for $\sigma(x) = 0.1$



Comparison of weights w.r.t. $\mu(x) - T$ for $\sigma(x) = 0.01$



Comparison of weights w.r.t. $\sigma(x)$ for $\mu(x) - T$



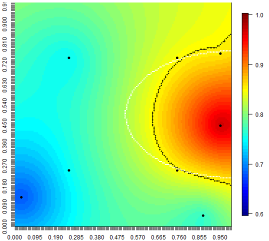
Interpretation of the weight $w_{\sigma_\epsilon^2}(x)$ for $\sigma_\varepsilon^2 = 0$

$$w_{\sigma_\epsilon^2=0}(x) = \lim_{a \to 0} \frac{1}{2a} \mathbb{P}\Big(y(x) \in [T - a; T + a]\Big)$$

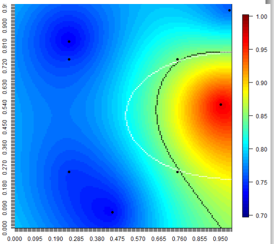
The additional weight σ_ε^2 is used to stabilize the weight for small values of $\sigma^2(x)$.

Example 1 : Comparison of 8-point designs

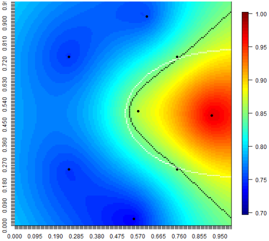
Map of μ
and level set
for MC^{thres}.



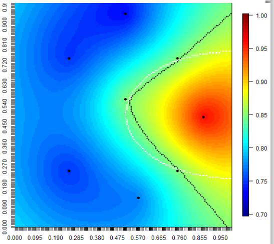
Map of μ
and level set
for $tMMSE$.



Map of μ
and level set
for IC^{thres}.

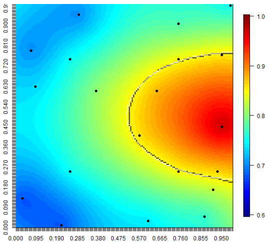


Map of μ
and level set
for $tIMSE$.

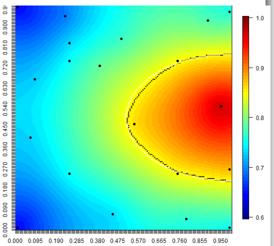


Example 1 : Comparison of 20-point designs

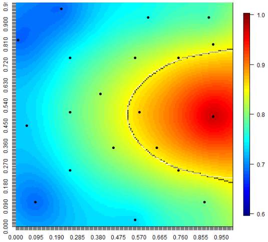
Map of μ and level set for MC^{thres}.



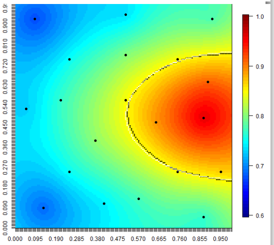
Map of μ and level set for $tMMSE$.



Map of μ and level set for IC^{thres}.



Map of μ and level set for $tIMSE$.

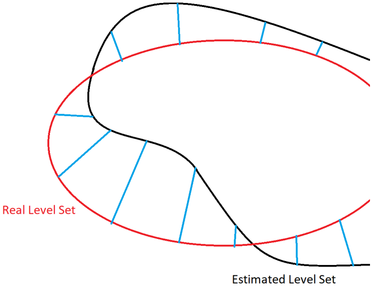


Quality score for level line detection

Q_{dist}-score

Average distance between the estimated level line and the true level line and conversely

$$Q_{\mathrm{dist}} = \left(Q_{\mathrm{dist}}^{\mathrm{real}} + Q_{\mathrm{dist}}^{\mathrm{est}}\right) \ / \ 2$$



- $Q_{\mathrm{dist}}^{\mathrm{est}}$ is the average distance between each point of $\mathcal{L}^{\mathrm{est}}$ and the closest point of $\mathcal{L}^{\mathrm{real}}$;
- $Q_{\mathrm{dist}}^{\mathrm{real}}$ is the average distance between each point of $\mathcal{L}^{\mathrm{real}}$ and the closest point of $\mathcal{L}^{\mathrm{est}}$.

Q_μ-score

$$Q_\mu = \left(Q_\mu^{\text{real}} + Q_\mu^{\text{est}}\right) \ / \ 2.$$

where

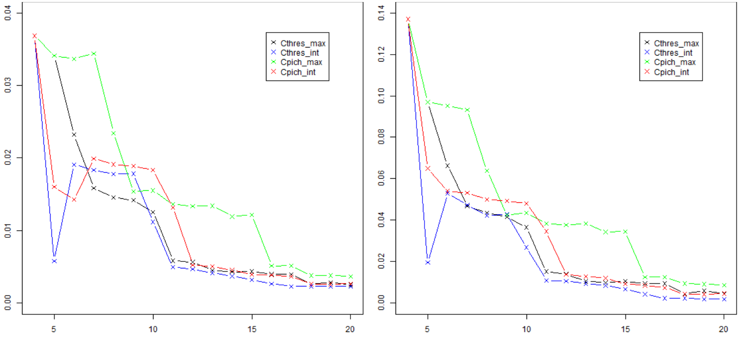
$$Q_\mu^{\text{est}} = \frac{1}{\#\{\mathcal{L}^{\text{est}}\}} \sum_{x \in \mathcal{L}^{\text{est}}} |y(x) - T|$$

and

$$Q_\mu^{\text{real}} = \frac{1}{\#\{\mathcal{L}^{\text{real}}\}} \sum_{x \in \mathcal{L}^{\text{real}}} |\widehat{y}(x) - T|$$

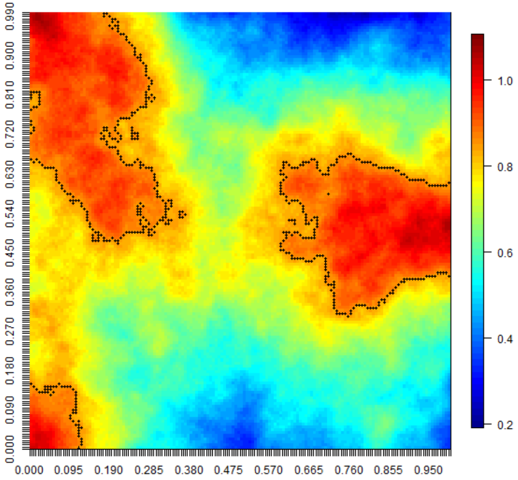
- Q_μ^{est} is the average discrepancy between the true value $y(x)$ and the threshold T on the estimated level line $\mathcal{L}^{\text{est}} = \{x \ / \ \widehat{y}(x) = T\}$.
- Q_μ^{real} is the average discrepancy between the estimated value $\widehat{y}(x)$ and the threshold T on the real level line $\mathcal{L}^{\text{real}} = \{x \ / \ y(x) = T\}$.

Evolution of Q_μ and Q_{dist} scores from 4 to 20-point designs.



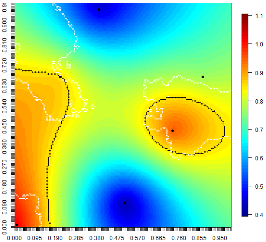
MC_T^{thres} / IC_T^{thres} / $tMMSE$ / $tIMSE$

A MORE COMPLEX EXAMPLE WITH DISCONNECTED AREAS

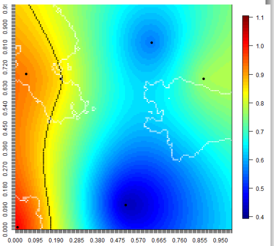


Comparison of 6-point designs

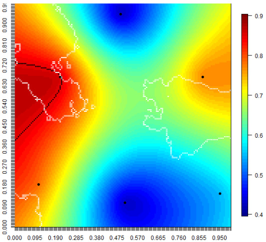
Map of μ and level set for MC^{thres}.



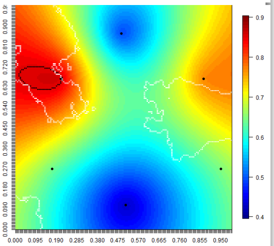
Map of μ and level set for $tMMSE$.



Map of μ and level set for IC^{thres}.

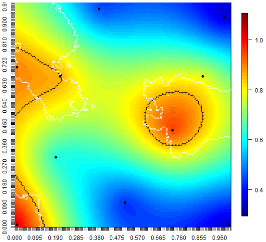


Map of μ and level set for $tIMSE$.

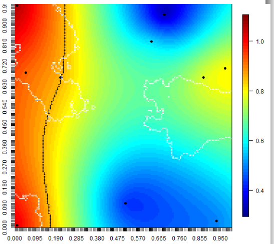


Comparison of 10-point designs

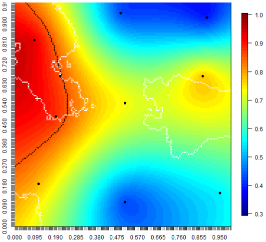
Map of μ and level set for MC^{thres}.



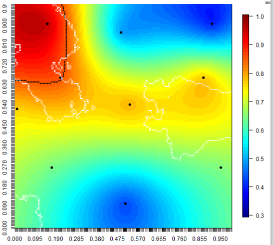
Map of μ and level set for $tMMSE$.



Map of μ and level set for IC^{thres}.

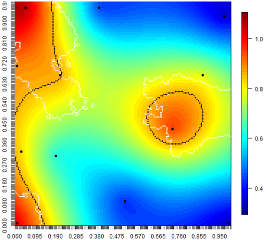


Map of μ and level set for $tIMSE$.

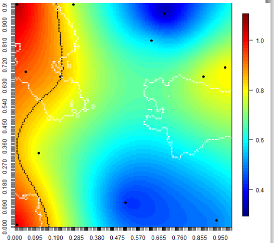


Comparison of 12-point designs

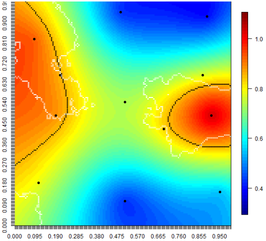
Map of μ and level set for MC^{thres}.



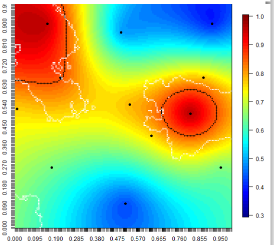
Map of μ and level set for $tMMSE$.



Map of μ and level set for IC^{thres}.

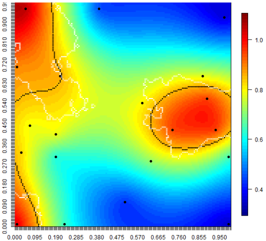


Map of μ and level set for $tIMSE$.

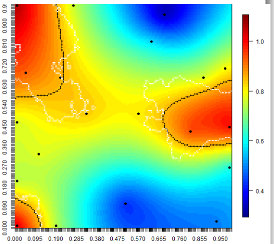


Comparison of 20-point designs

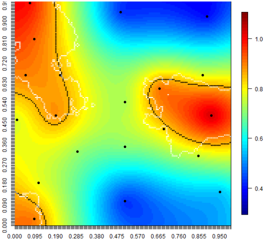
Map of μ and level set for MC^{thres}.



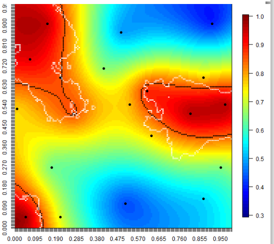
Map of μ and level set for $tMMSE$.



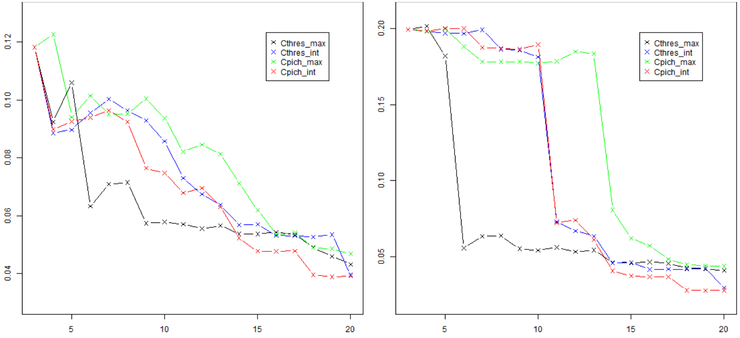
Map of μ and level set for IC^{thres}.



Map of μ and level set for $tIMSE$.



Evolution of Q_μ and Q_{dist} scores from 4 to 20-point designs.



MC_T^{thres} / IC_T^{thres} / $tMMSE$ / $tIMSE$

Conclusion and further work

New criteria

- C^{exc} : To minimize the estimation error where values exceed a threshold of interest.
- C^{high} : To minimize the estimation error where high values are expected.

Exchange algorithm alternatives

- Different exchange methods ;
- Simulated annealing optimization ;

$\longrightarrow$ No significant gain.

Conclusion and further work

New criteria

- C^{exc} : To minimize the estimation error where values exceed a threshold of interest.
- C^{high} : To minimize the estimation error where high values are expected.

Exchange algorithm alternatives

- Different exchange methods ;
- Simulated annealing optimization ;

$\longrightarrow$ No significant gain.

Conclusions and perspectives

Conclusion

- Our max criterion MC^{thres} perform well with few points or when few knowledge is available.
- Both integrated criteria IC^{thres} and tIMSE perform equivalently, with difficulties to detect new components of the area of interest.
- The max criterion tMMSE perform poorly.

Perspective

- Use successively MC^{thres} then IC^{thres}.
- Use the date to re-estimate the correlation structure.