Field Parametrization for Bayesian Inference

POLETTE Nadège, *CEA DAM DIF/Mines Paris PSL*

Supervisors:

SOCHALA Pierre, *CEA DAM DIF*

GESRET Alexandrine, *Mines Paris PSL*

LE MAÎTRE Olivier, *CMAP, CNRS*

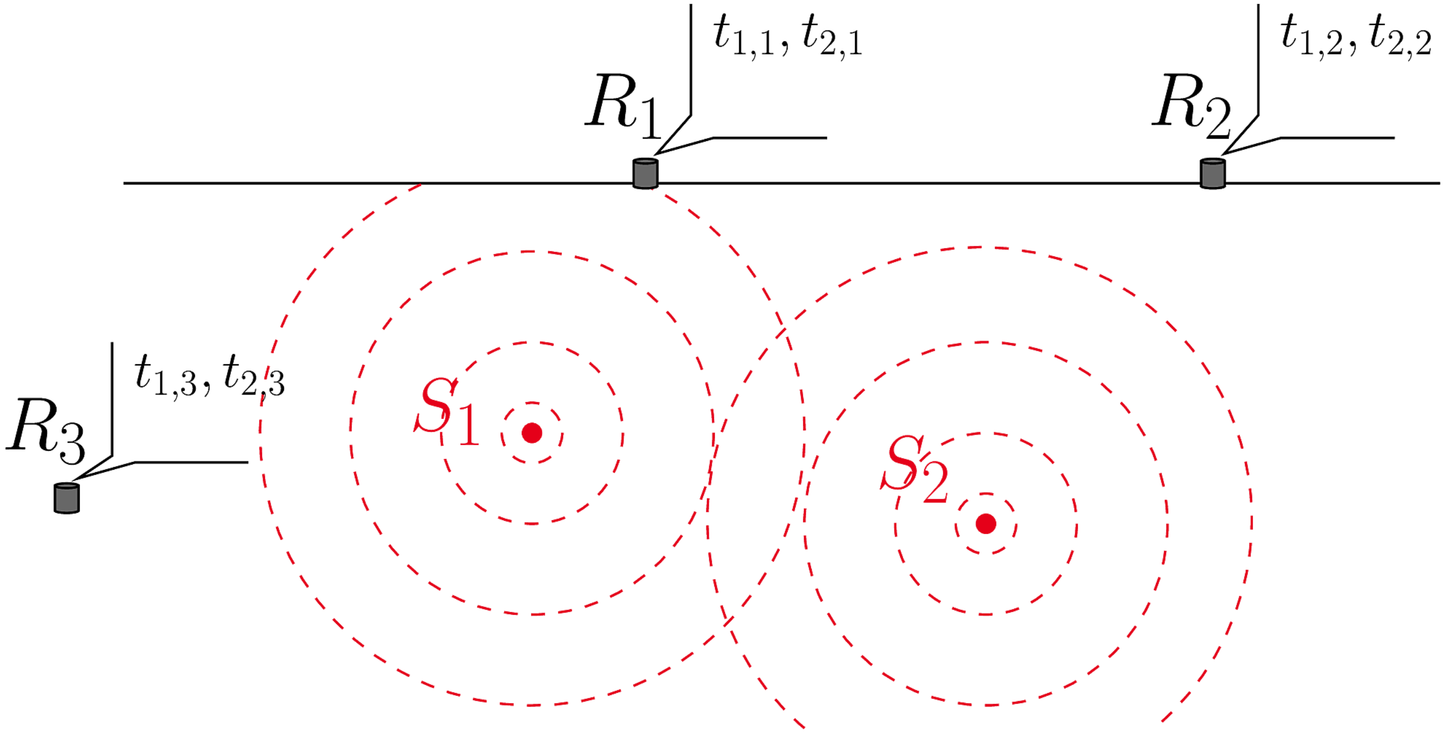
Table of contents

1. Context: Detection and analysis of seismic events

Global scale
- International treaties (CTBT, NTP)
- Environment monitoring (IMS)

Regional scale
- Tsunami and earthquake alerts
- Risk prevention

Local scale
- Subsurface knowledge
- Exploitation

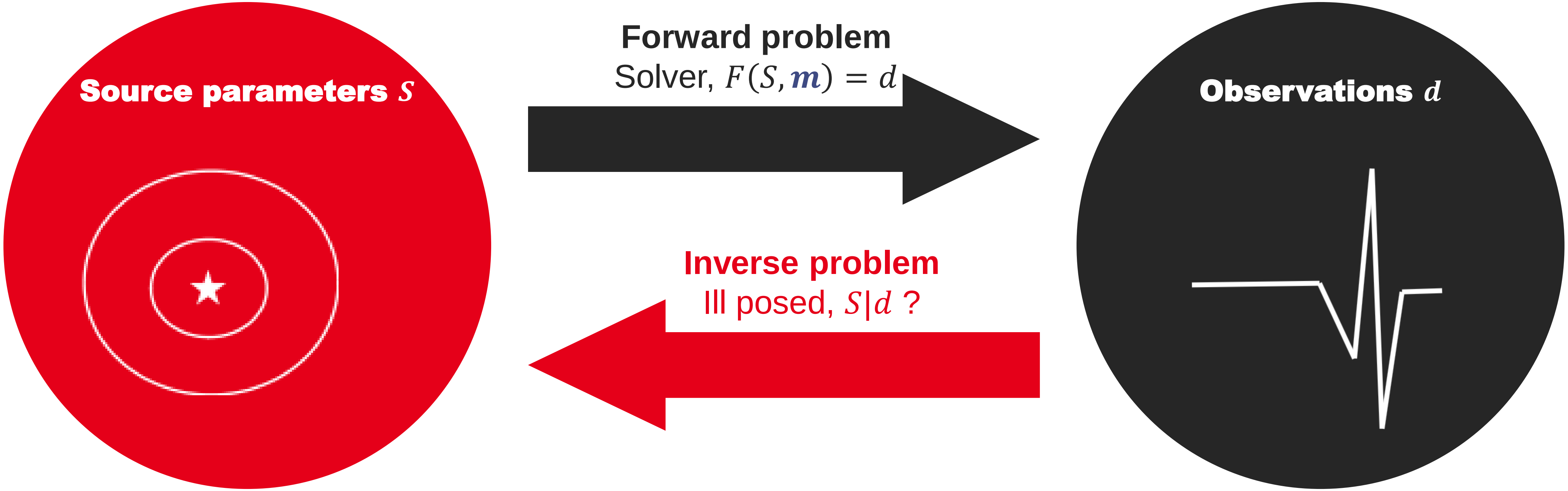


$$F(S) = d$$

F: forward model
S: source parameters
d: data

Objective: retrieve S from d
- fast
- with accuracy
- with uncertainties

1. Context: Inverse problem

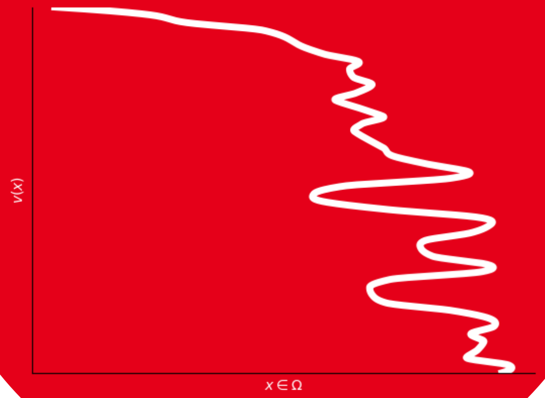


Uncertainty sources: observations, physical model, **model parameters**, ...

[Tarantola, 2005]

1. Context: Inverse problem

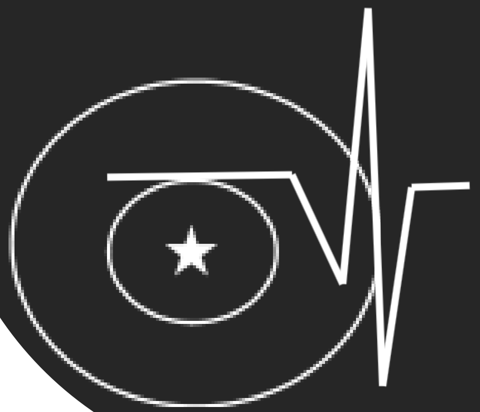
Model parameters m



Forward problem
Solver, $F(S, \boldsymbol{m}) = d$

Inverse problem
Ill posed, $m|d, S$?

Observations d,
Source parameters S



Objective: improve uncertainty quantification of **model parameters**

[Tarantola, 2005]

1. Context: Bayesian inference and Markov Chain Monte Carlo

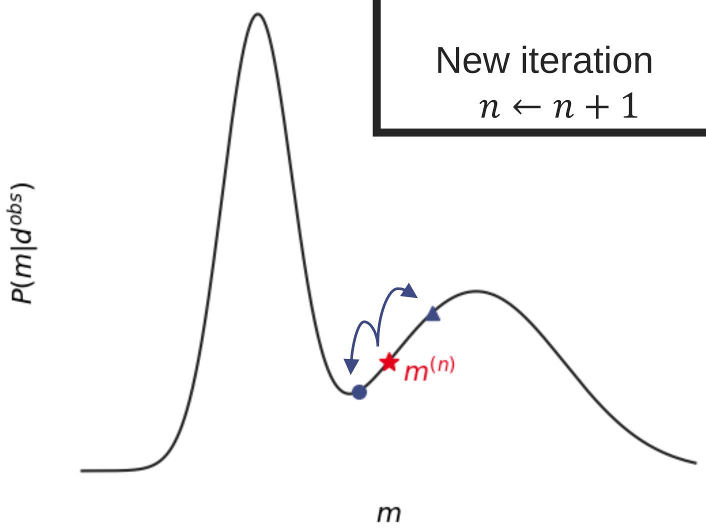
Forward model
Solver $F(m; S) = d$

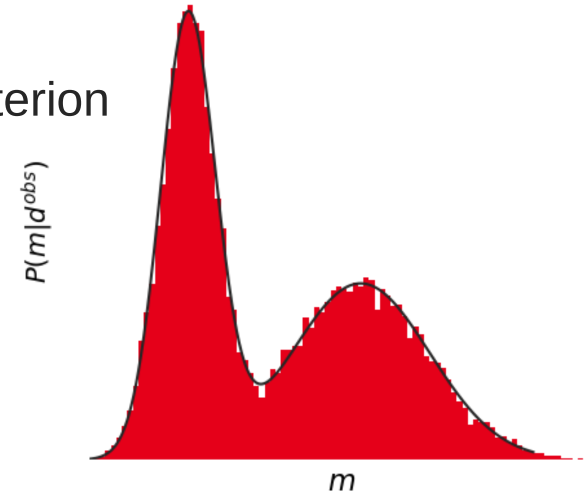
Proposition $m^{(n)}$ → Comparison d vs d^{obs} (data)

Bayes' rule : $P(m|d^{obs}) \propto L(d^{obs}|m)P(m)$

Accept/Reject criterion

New iteration
$n \leftarrow n + 1$



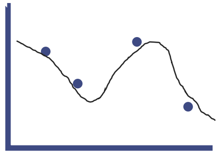


[Sivia,Skilling, 2006]
[Doucet et al., 2013]

1. Context: Bayesian inference

Surrogate model

Polynomial chaos $\widetilde{PC}(m) = \sum_k f_k \Psi_k(m) = d$



Proposition $m^{(n)}$

m infinite dimensional

→ dimension reduction

Parametrization:

- 'small' dimension
- fixed dimension
- allowing for various shapes

Comparison

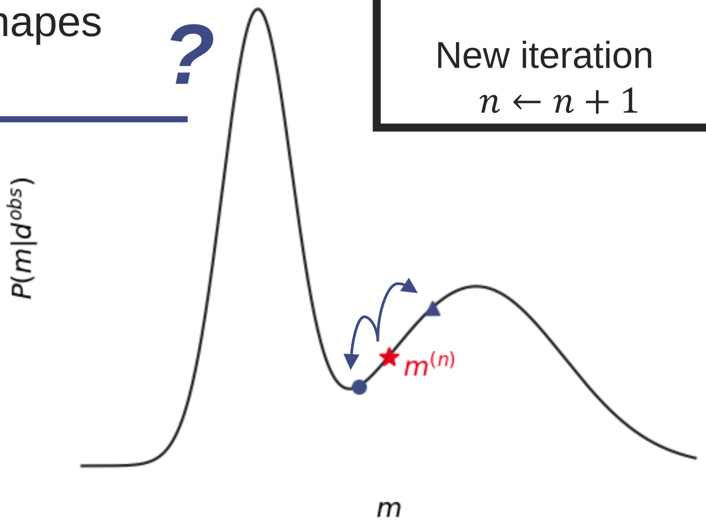
d vs d^{obs} (data)

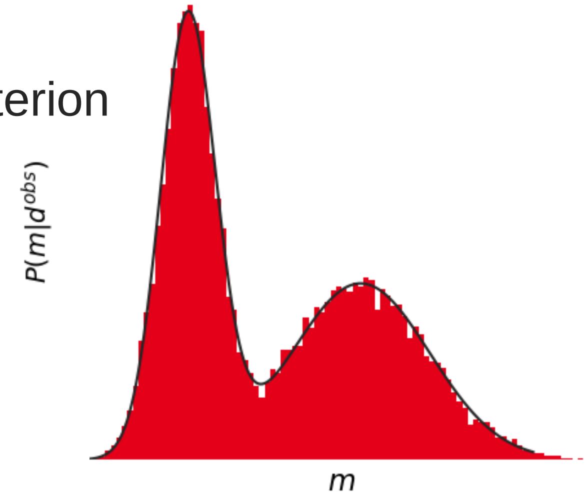
Bayes' rule : $P(m|d^{obs}) \propto L(d^{obs}|m)P(m)$

Accept/Reject criterion

New iteration

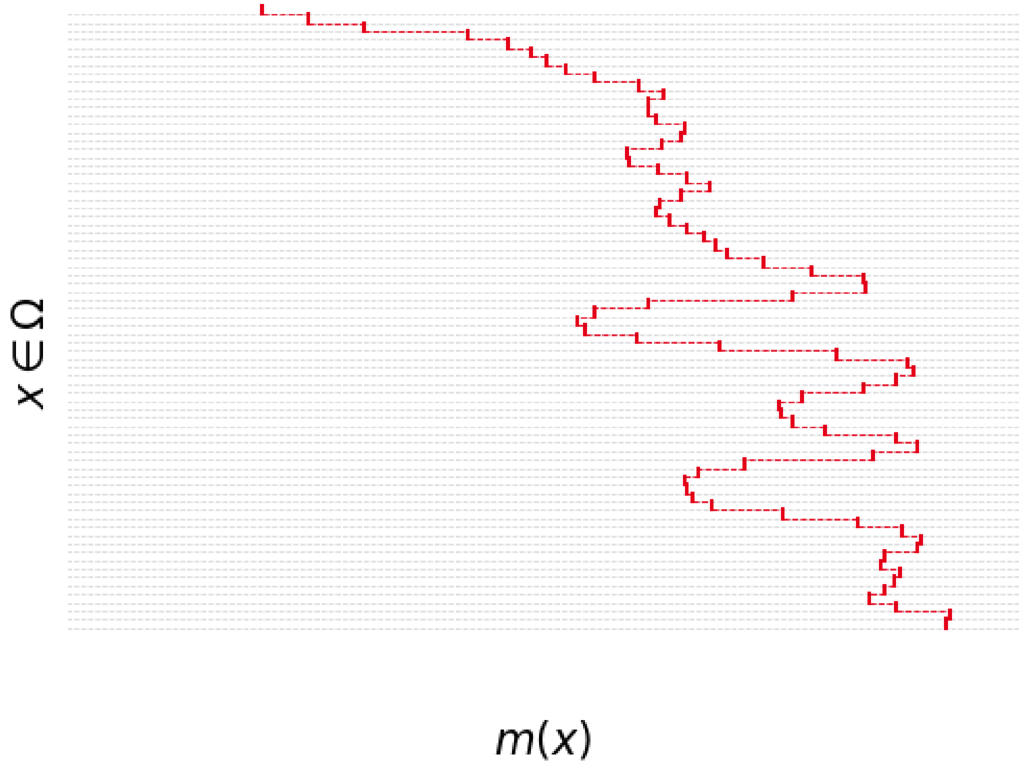
$n \leftarrow n+1$





[Wiener, 1938]
[Ghanem, Spanos, 1991]
[Xiu, Karniadakis, 2002]

2. Field parametrization: Spatial mesh

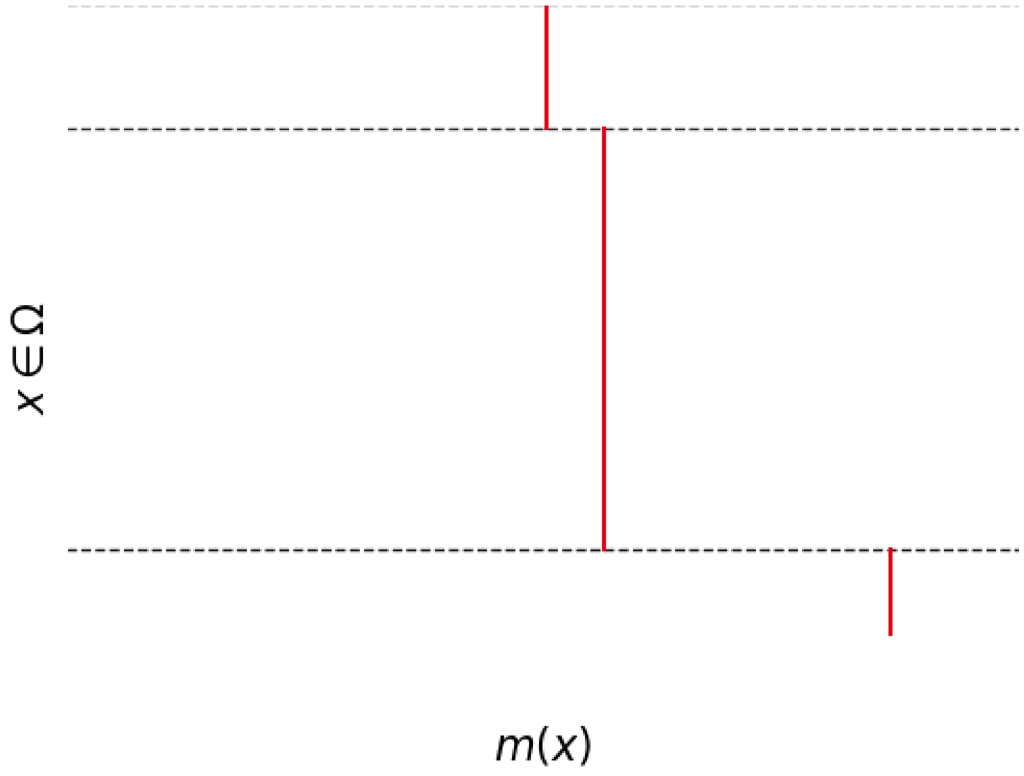


$\forall x \ \in [x_{i,} x_{i+1}], \ m(x) = m_i$

Parameters: $\{m_i\}_{1 \leq i \leq N_{meshes}}$

✖ 'small' dimension
OK fixed dimension
OK allowing for various shapes

2. Field parametrization: Layered velocity model



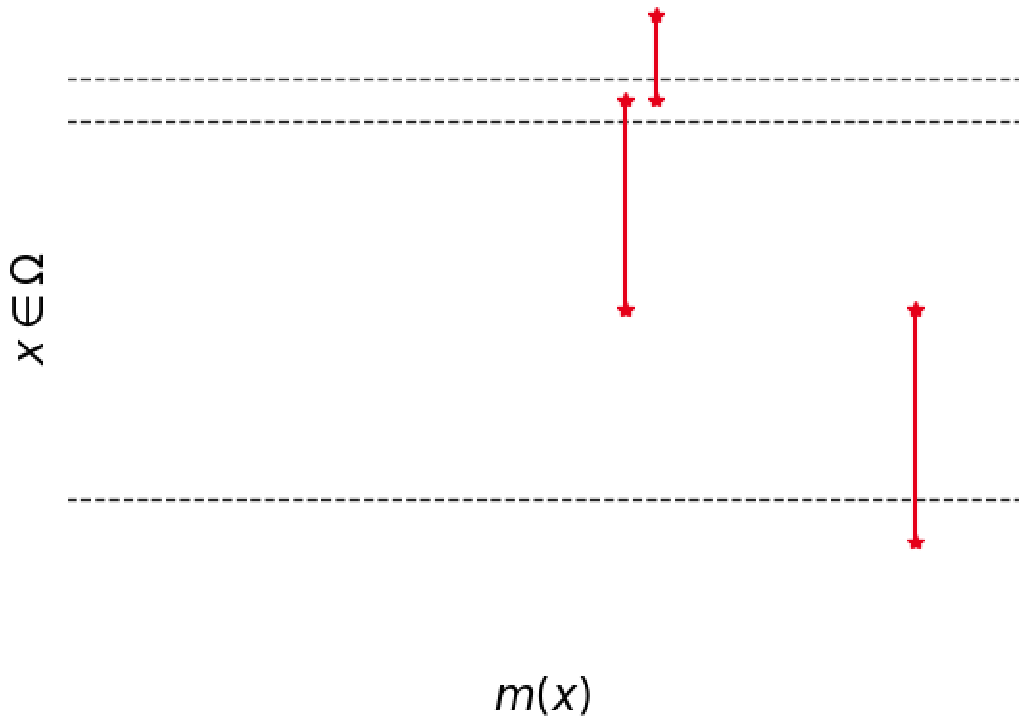
$$\forall x \in [x_i, x_{i+1}],\ m(x) = m_i$$

Parameters: $\{m_i, z_i\}_{1 \le i \le N_{layers}}$

OK 'small' dimension
OK fixed dimension
✖ allowing for various shapes

[Sochala et al., 2021]

2. Field parametrization: Voronoi tesselation



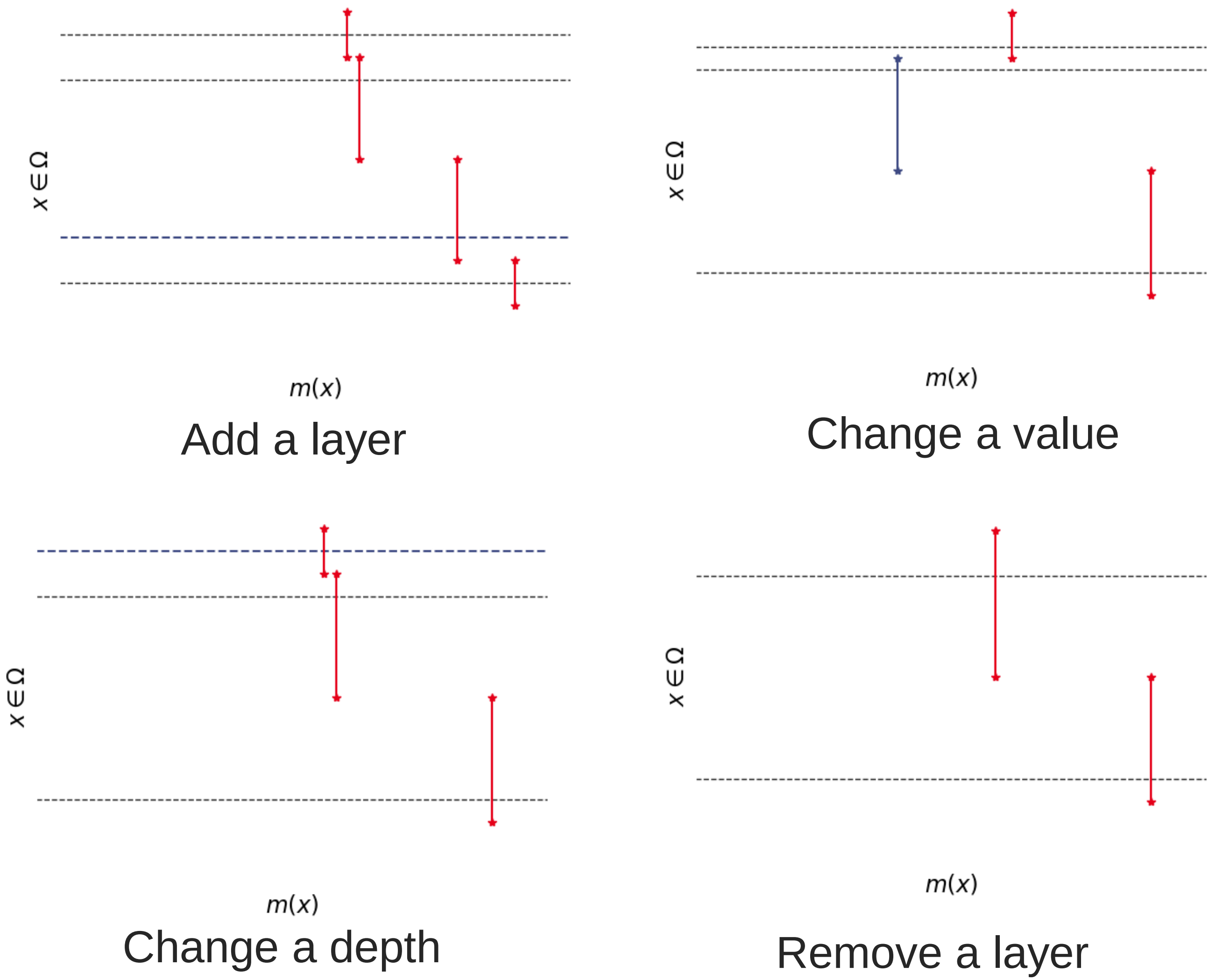
$$\forall x \in V(z_i),\ m(x) = m_i$$

Parameters: $\{m_i, z_i\}_{1 \le i \le n_c} \cup n_c$

OK 'small' dimension
✖ fixed dimension
OK allowing for various shapes

[Bodin et al., 2012]
[Piana Agostinetti et al., 2015]
[Belhadj et al., 2018]

2. Field parametrization: Voronoi tesselation



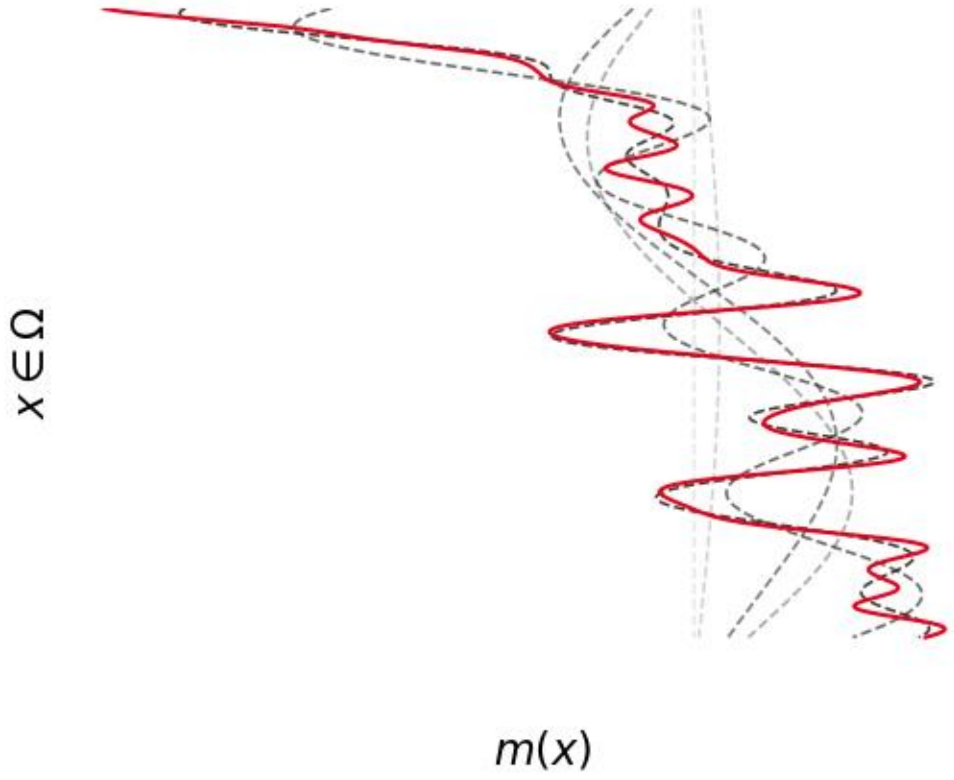
$$\forall x \in V(z_i),\ m(x) = m_i$$

Parameters: $\{m_i, z_i\}_{1 \leq i \leq n_c} \cup n_c$

OK 'small' dimension
✖ fixed dimension
OK allowing for various shapes

[Bodin et al., 2012]
[Piana Agostinetti et al., 2015]
[Belhadj et al., 2018]

2. Field parametrization: Modal representation



$$m(x) = \sum_{i=1}^{r} U_i(x) w_i$$

Parameters: $\{w_i\}_{1 \leq i \leq r}$

OK ‘small’ dimension
OK fixed dimension
~ allowing for various shapes

[Marzouk, Najm, 2009]

2. Field parametrization: Karhunen-Loève decomposition

Assuming m is the realization of a random process with autocovariance function k

$$m(x) = \sum_{i=1}^{r} \sqrt{\lambda_i} u_i(x) \eta_i$$

$(\lambda_i, u_i)_{1 \leq i \leq r}$ eigenelements of k: $\langle k(x,\cdot), u_i \rangle = \int_\Omega k(x,y) u_i(x) dx = \lambda_i u_i(y)$

The decomposition is bi-orthonormal:

- $\langle u_i, u_j \rangle_\Omega = \delta_{i,j}$
- $E(\eta_i) = 0$, and $E(\eta_i \eta_j) = \delta_{i,j}$

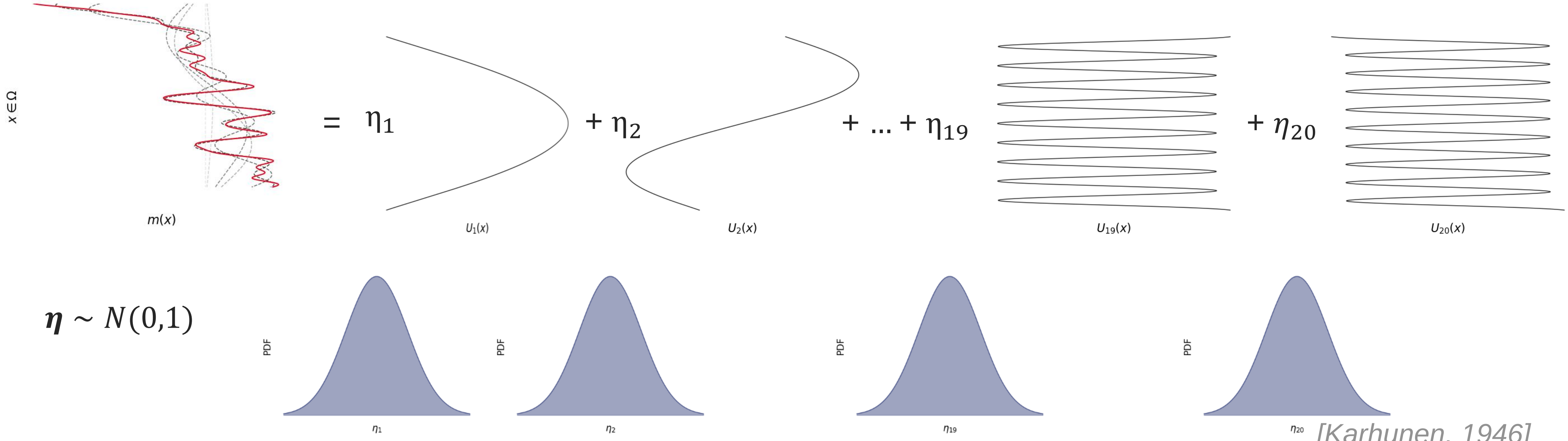
In the case of a Gaussian process, we have $\boldsymbol{\eta} \sim N(0,1)$

[Karhunen, 1946]
[Loeve, 1977]
[Marzouk, Najm, 2009]

2. Field parametrization: Karhunen-Loève decomposition

Assuming m is the realization of a random process with autocovariance function k

$$m(x) = \sum_{i=1}^{r} \sqrt{\lambda_i} u_i(x) \eta_i$$



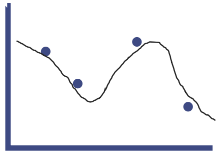
$\boldsymbol{\eta} \sim N(0,1)$

[Karhunen, 1946]
[Loeve, 1977]
[Marzouk, Najm, 2009]

2. Field parametrization: Karhunen-Loève decomposition

Surrogate model

Polynomial chaos $\widetilde{PC}(\boldsymbol{\eta}) = \sum_k f_k \Psi_k(\boldsymbol{\eta}) = d$



Proposition $\boldsymbol{\eta}^{(n)}$

OK 'small' dimension

OK fixed dimension

~ allowing for various shapes

Comparison

d vs d^{obs} (data)

Bayes' rule : $\boldsymbol{P(\eta|d^{obs}) \propto L(d^{obs}|\eta)P(\eta)}$

$\boldsymbol{\eta} \sim N(0,1)$

New iteration

$n \leftarrow n+1$

Accept/Reject criterion

Problem : how to choose $\boldsymbol{k}$?

[Marzouk, Najm 2009]

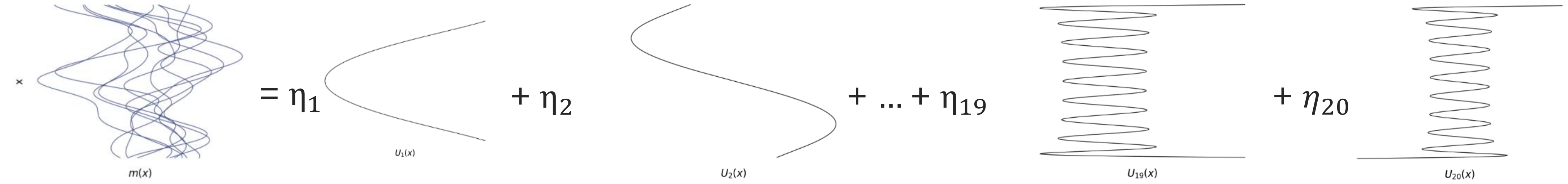
3. Change of measure: Objective

Assuming k is a Gaussian autocovariance function: $k(x,y) = A \exp\left(\frac{-\|x-y\|^2}{2l^2}\right)$

It depends on two hyperparameters : $\boldsymbol{q} = \{A, l\}$

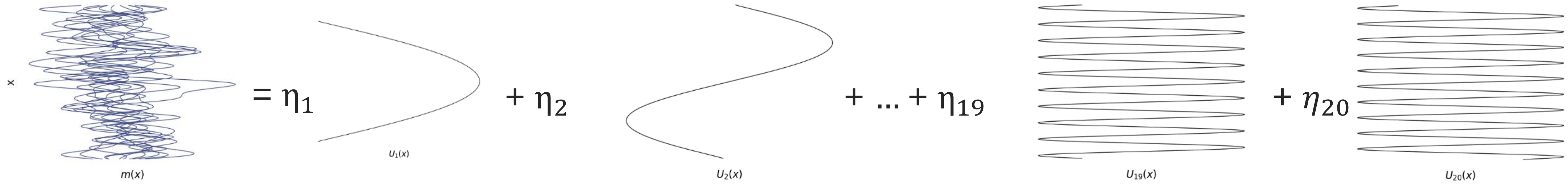
$$m(x) = \sum^{r} \sqrt{\lambda_i(\boldsymbol{q})}\, u_i(x,\boldsymbol{q})\eta_i$$

Large l



$\boldsymbol{\eta} \sim N(0,1)$

Small l



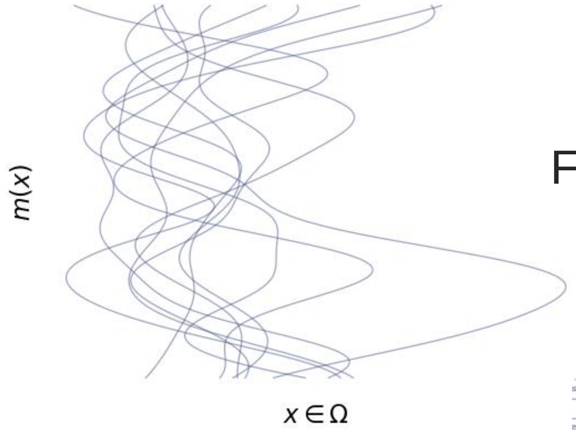
3. Change of measure: Objective

Assuming k is a Gaussian autocovariance function: $k(x,y) = A \exp\left(\frac{-\|x-y\|^2}{2l^2}\right)$

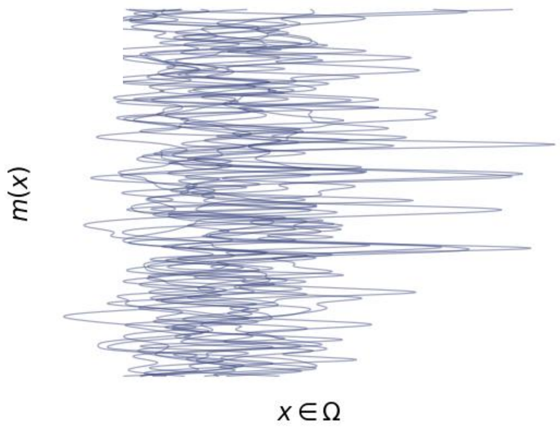
It depends on two hyperparameters : $\boldsymbol{q} = \{A, l\}$

$$m(x) = \sum_{i=1}^{r} \sqrt{\lambda_i(\boldsymbol{q})} u_i(x, \boldsymbol{q}) \eta_i$$

Field realizations



Large l



Small l

- Hyperparameters are determined *a priori*

→ expert judgement, MSE, LOOCV…

Overconfidence risk

[Rasmussen, Williams, 2015]

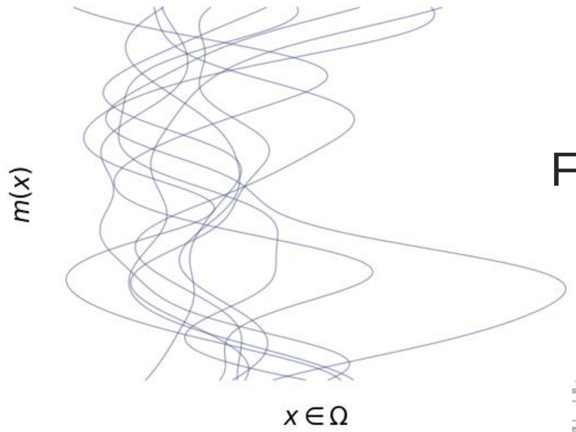
3. Change of measure: Objective

Assuming k is a Gaussian autocovariance function: $k(x,y) = A \exp\left(\frac{-\|x-y\|^2}{2l^2}\right)$

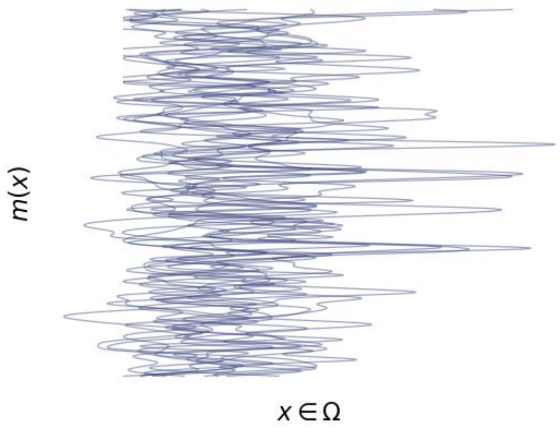
It depends on two hyperparameters : $\boldsymbol{q} = \{A, l\}$

$$m(x) = \sum_{i=1}^{r} \sqrt{\lambda_i(\boldsymbol{q})}\, u_i(x, \boldsymbol{q})\, \eta_i$$

Field realizations



Large l



Small l

- Hyperparameters are determined *a priori*
  → expert judgement, MSE, LOOCV…
  Overconfidence risk

- Hyperparameters are inferred during the procedure
  → Bayes' rule : $\boldsymbol{P}(\boldsymbol{\eta}, \boldsymbol{q} | \boldsymbol{d}^{obs}) \propto \boldsymbol{L}(\boldsymbol{d}^{obs} | \boldsymbol{\eta}, \boldsymbol{q}) \boldsymbol{P}(\boldsymbol{\eta}, \boldsymbol{q})$
  Expensive

[Rasmussen, Williams, 2015]
[Tagade, Choi, 2014]
[Sraj et al., 2016]

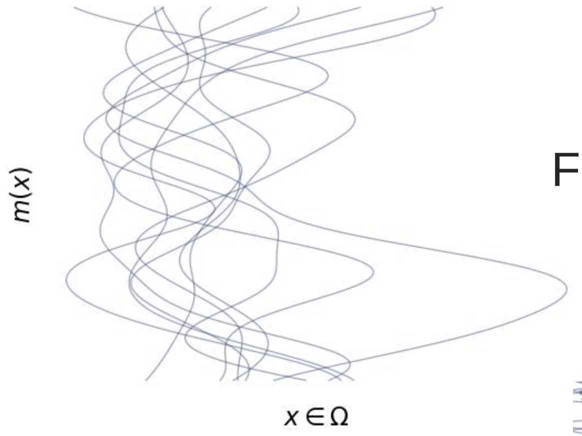
3. Change of measure: Objective

Assuming k is a Gaussian autocovariance function: $k(x,y) = A \exp\left(\frac{-\|x-y\|^2}{2l^2}\right)$

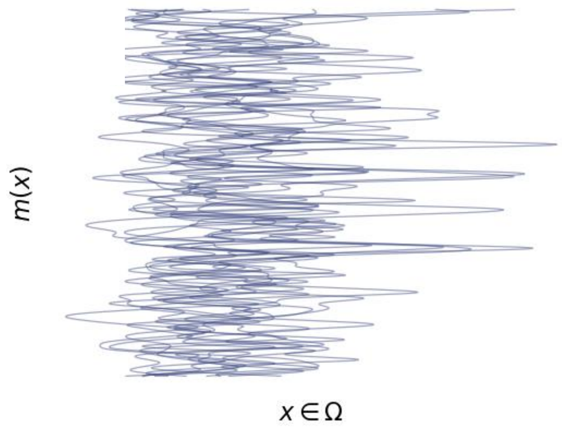
It depends on two hyperparameters : $\boldsymbol{q} = \{A, l\}$

$$m(x) = \sum_{i=1}^{r} \sqrt{\lambda_i(\boldsymbol{q})}\, u_i(x,\boldsymbol{q})\, \eta_i$$

Field realizations



Large l



Small l

- Hyperparameters are determined *a priori*

→ expert judgement, MSE, LOOCV…

Overconfidence risk

- Hyperparameters are inferred during the procedure

→ Bayes' rule : $\boldsymbol{P}(\boldsymbol{\eta},\boldsymbol{q}|\boldsymbol{d}^{obs}) \propto \boldsymbol{L}(\boldsymbol{d}^{obs}|\boldsymbol{\eta},\boldsymbol{q})\boldsymbol{P}(\boldsymbol{\eta},\boldsymbol{q})$

Expensive

Objective : develop a cheap method to take into account hyperparameters

[Rasmussen, Williams, 2015]
[Tagade, Choi, 2014]
[Sraj et al., 2016]

3. Change of measure: Reference basis

We introduce a reference kernel $\bar{k} = \int_H k(\cdot,\cdot,\boldsymbol{q})d\boldsymbol{q}$.

The reference basis $(\bar{u}_i, \bar{\lambda}_i)_{1\leq i\leq r}$ are the eigenelements of $\bar{k}$: $\int_\Omega \bar{k}(x,y)\bar{u}_i(x)dx = \bar{\lambda}_i\bar{u}_i(y)$.

The field decomposition writes $$m(x) = \sum_{i=1}^{r} \sqrt{\bar{\lambda}_i}\bar{u}_i(x)\xi_i$$

Hierarchical Bayes formulation:

→ Bayes' rule : $\boldsymbol{P}(\boldsymbol{\xi},\boldsymbol{q}|\boldsymbol{d}^{obs}) \propto \boldsymbol{L}(\boldsymbol{d}^{obs}|\boldsymbol{\xi})\boldsymbol{P}(\boldsymbol{\xi},\boldsymbol{q}) = \boldsymbol{L}(\boldsymbol{d}^{obs}|\boldsymbol{\xi})\boldsymbol{P}(\boldsymbol{\xi}|\boldsymbol{q})\boldsymbol{P}(\boldsymbol{q})$

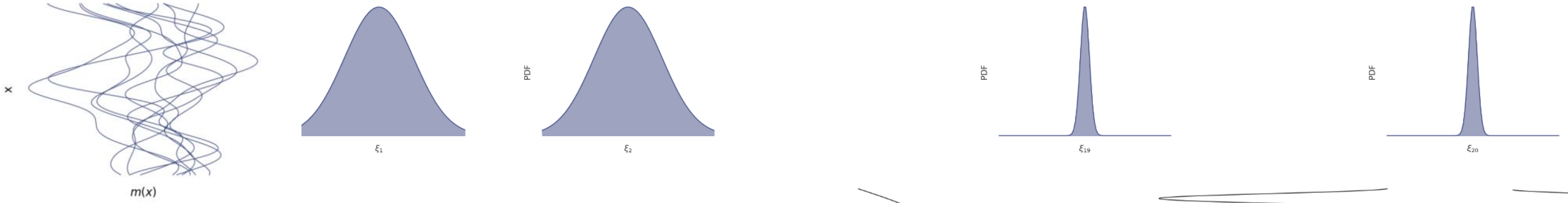
⇒ The $\boldsymbol{q}$-dependency is transferred to the prior law of the coordinates $\boldsymbol{\xi}$

[Sraj et al., 2016]
[Polette et al., 2024]

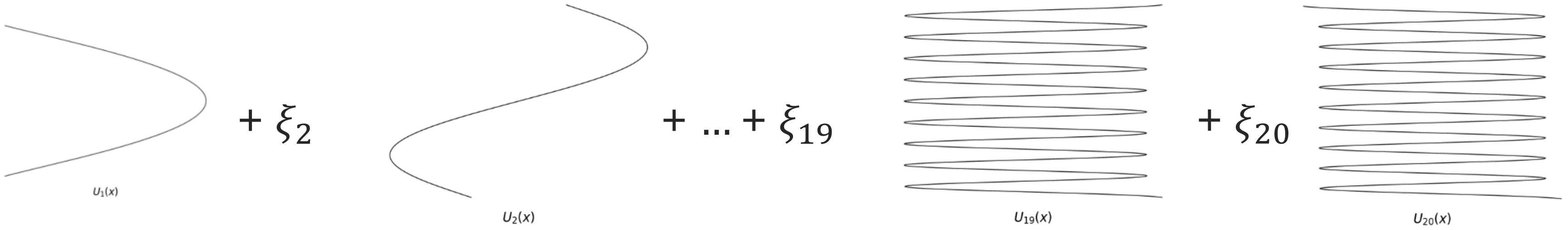
3. Change of measure: Reference basis

The field decomposition writes $$m(x) = \sum_{i=1}^{r} \sqrt{\bar{\lambda}_i}\, \bar{u}_i(x)\, \xi_i\,, \qquad \boldsymbol{\xi} \sim P(\boldsymbol{\xi}|\boldsymbol{q})$$

Large l



Reference basis $= \xi_1$ $+ \xi_2$ $+ \ldots + \xi_{19}$ $+ \xi_{20}$



Small l

How is defined $P(\boldsymbol{\xi}|\boldsymbol{q})$?

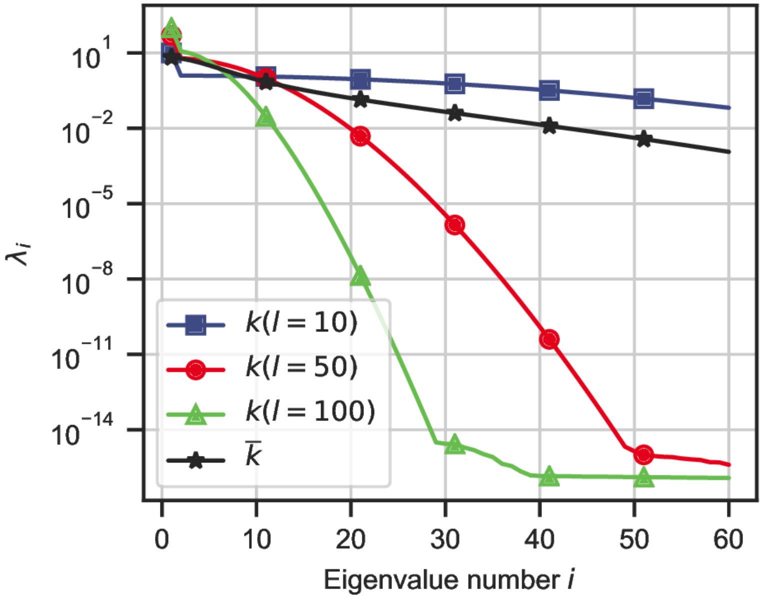
[Sraj et al., 2016]
[Polette et al., 2024]

3. Change of measure: Formulation

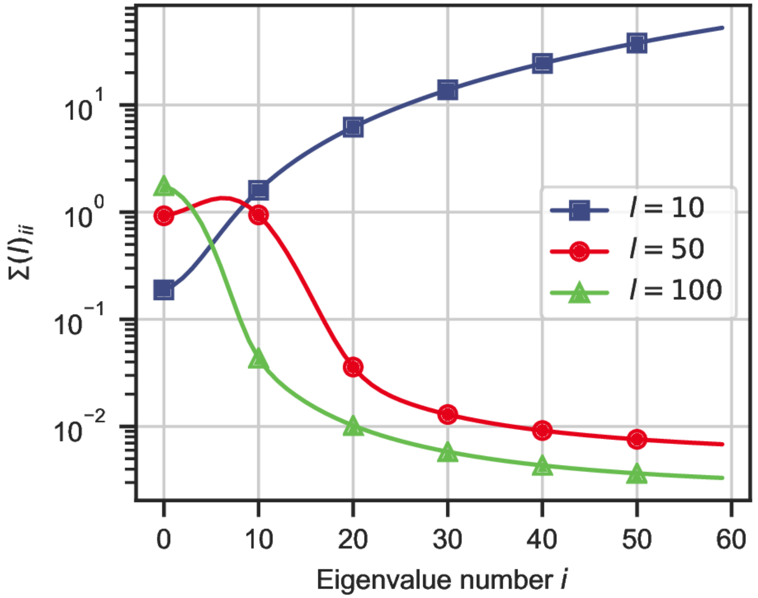
Prior law of the reference coordinates according to the hyperparameters:

$$m(x) = \sum_{i=1}^{r} \sqrt{\bar{\lambda}_i} \bar{u}_i(x)\xi_i, \ \boldsymbol{\xi} \sim N(0, \Sigma(\boldsymbol{q})) \ with \ \Sigma(\boldsymbol{q})_{i,j} = (\bar{\lambda}_i \bar{\lambda}_j)^{-1/2} \left\langle \langle k(\cdot,\cdot,\boldsymbol{q}), \bar{u}_i \rangle, \bar{u}_j \right\rangle$$

$\Sigma(\boldsymbol{q})$ is the double projection of the $\boldsymbol{q}$-dependent kernel on the reference basis



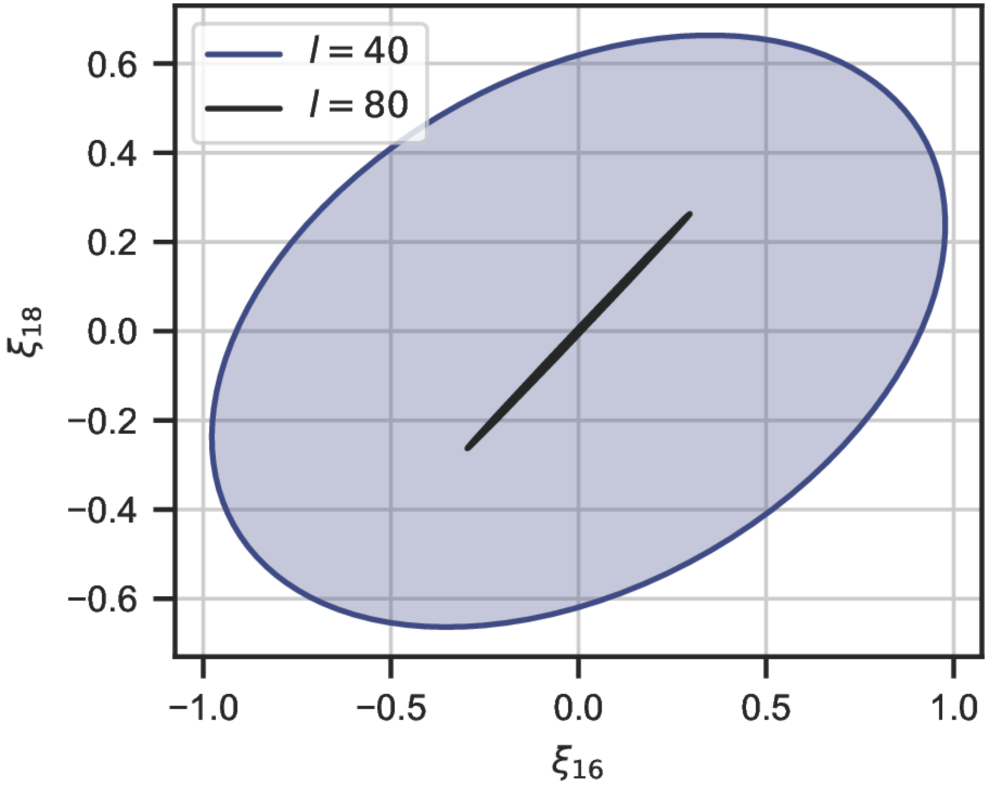
Eigenvalues decay according to the basis



Variance of $\boldsymbol{\xi}$ according to the hyperparameters

3. Change of measure: Sampling

Hierarchical sampling: $\boldsymbol{\xi} \sim N\big(0, \Sigma(\boldsymbol{q})\big)$, the prior distribution of $\boldsymbol{\xi}$ can be highly sensitive to $\boldsymbol{q}$



$\Sigma(\boldsymbol{q})$ Covariance projected on (ξ_8, ξ_6) space

Introduction of an auxiliary variable $\bar{\boldsymbol{\zeta}}$ whose prior law does not depend on hyperparameters

- Sample $\bar{\boldsymbol{\zeta}} \sim N(0,1)$, $\boldsymbol{q}$
- Compute $\boldsymbol{\xi} \sim N(0, \Sigma(\boldsymbol{q}))$ from $(\bar{\boldsymbol{\zeta}}, \boldsymbol{q})$ sample:

$$\boldsymbol{\xi} = \Sigma(\boldsymbol{q})^{1/2} \bar{\boldsymbol{\zeta}}$$

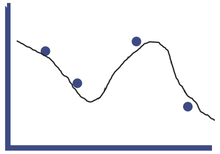
The proposition is not symmetric anymore, the ratio of the transition probabilities become

$$\frac{p(\boldsymbol{\xi}^{(n)}, \boldsymbol{q}^{(n)} | \boldsymbol{\xi}^{\star}, \boldsymbol{q}^{\star})}{p(\boldsymbol{\xi}^{\star}, \boldsymbol{q}^{\star} | \boldsymbol{\xi}^{(n)}, \boldsymbol{q}^{(n)})} = \left(\frac{\det\big(\Sigma(\boldsymbol{q}^{\star})\big)}{\det(\Sigma(\boldsymbol{q}^{(n)}))}\right)^{1/2}$$

[Betancourt, Girolami, 2013]

3. Change of measure: Summary

$$m(x) = \sum_{i=1}^{r} \sqrt{\bar{\lambda}_i}\,\bar{u}_i(x)\xi_i \text{, with } \boldsymbol{\xi} \sim N(0, \Sigma(\boldsymbol{q}))$$

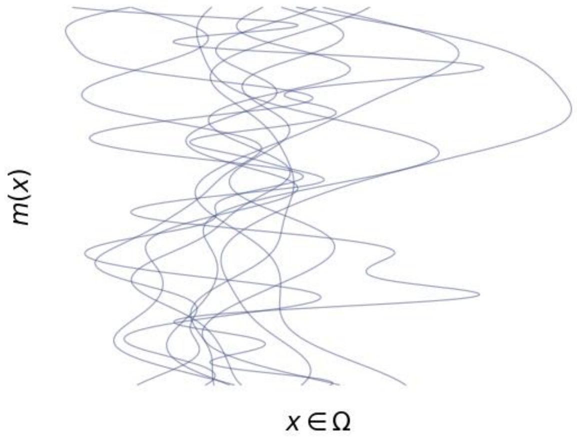


Surrogate model

Polynomial chaos $\widetilde{PC}(\boldsymbol{\xi}) = \sum_k f_k \Psi_k(\boldsymbol{\xi}) = d$

Proposition $\xi^{(n)}, \boldsymbol{q}^{(n)}$

OK 'small' dimension
OK fixed dimension
OK allowing for various shapes



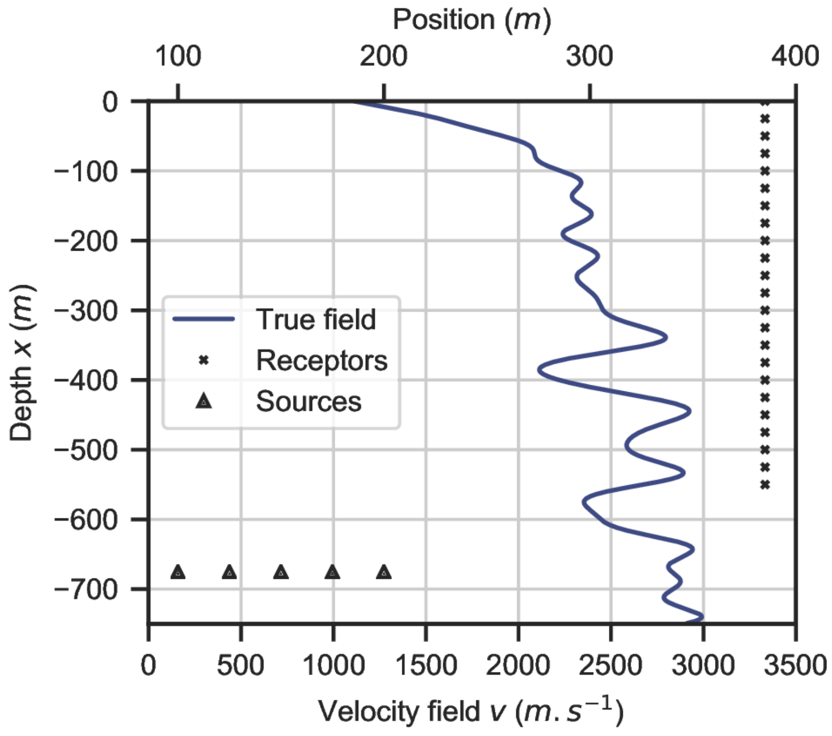
Comparison
d vs d^{obs} (data)

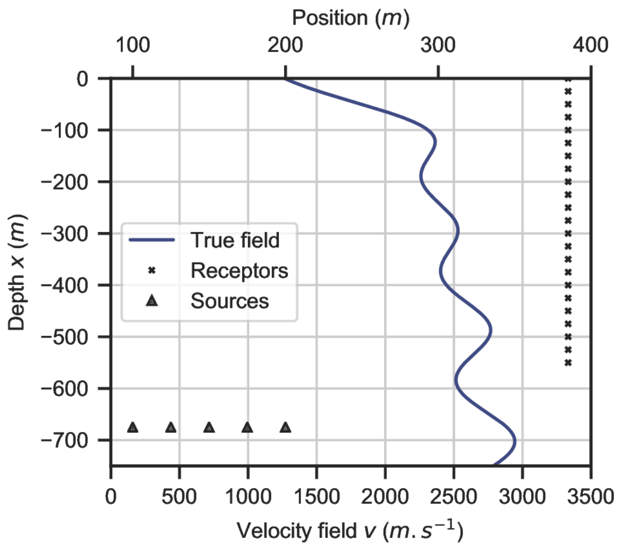
Bayes' rule : $\boldsymbol{P(\xi, q | d^{obs}) \propto L(d^{obs} | \xi) P(\xi | q) P(q)}$

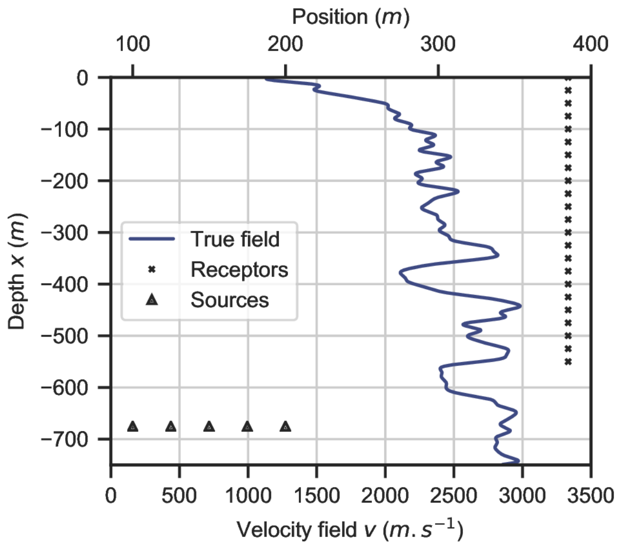
Accept/Reject criterion

New iteration
$n \leftarrow n + 1$

4. Results: Case presentation







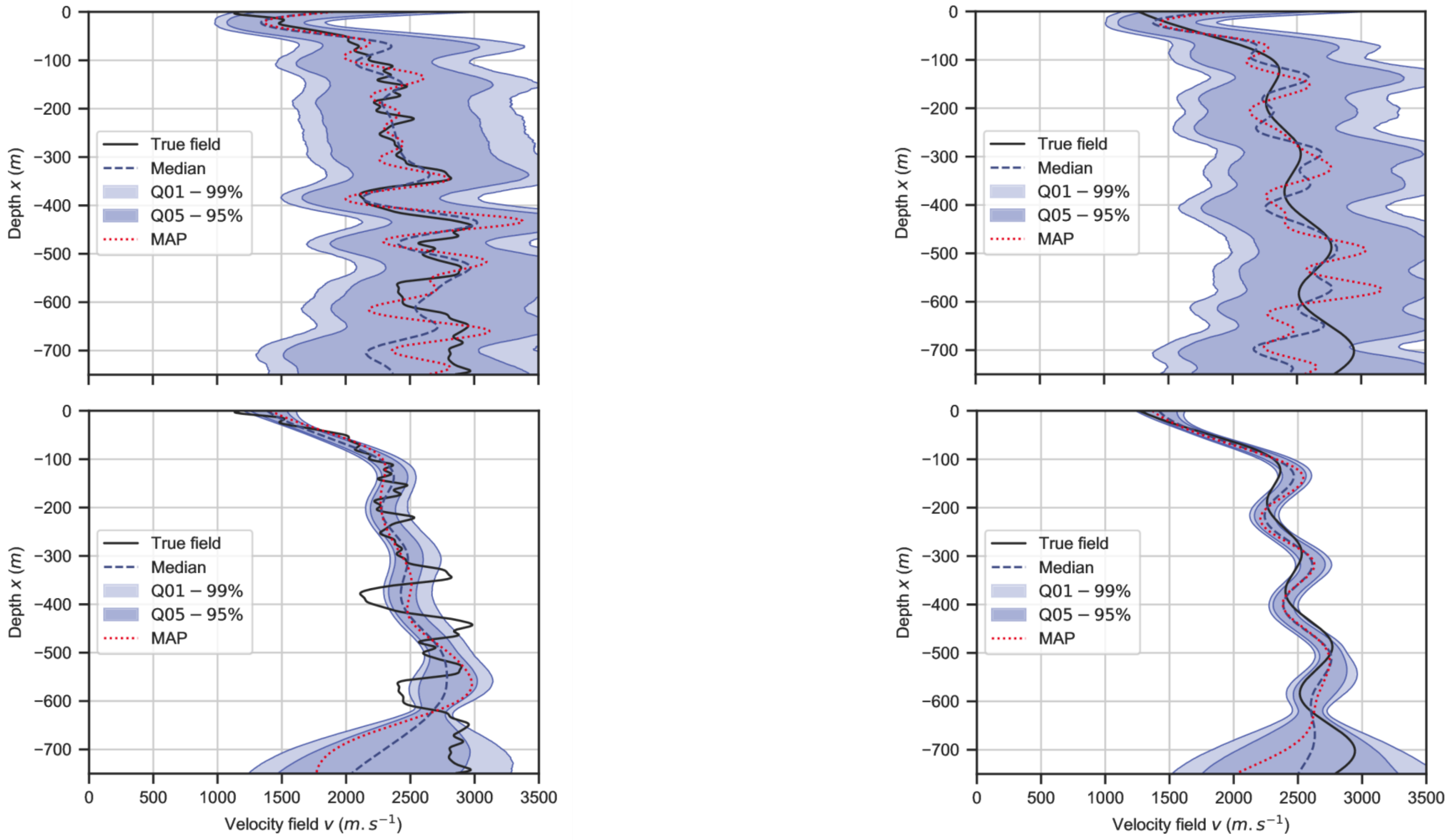
- Based on a realistic velocity model (*Amoco Tulsa Research Lab*)
- 2D velocity field, varies only along depth
- $\Omega = [0,750]m$, 23 stations × 5 events, noise level $0.002s$
- Velocity field writes $v(x) = \exp(\mu + \sum_{i=1}^{r} \sqrt{\overline{\lambda}_i}\overline{u}_i(x)\xi_i)$
- $l \sim U(10,100)$, $A \sim IG(21,1)$, $r = 20$, $\mu \sim U(6.9,8.1)$

4. Results: with fixed hyperparameters

Small wavelength field | Large wavelength field

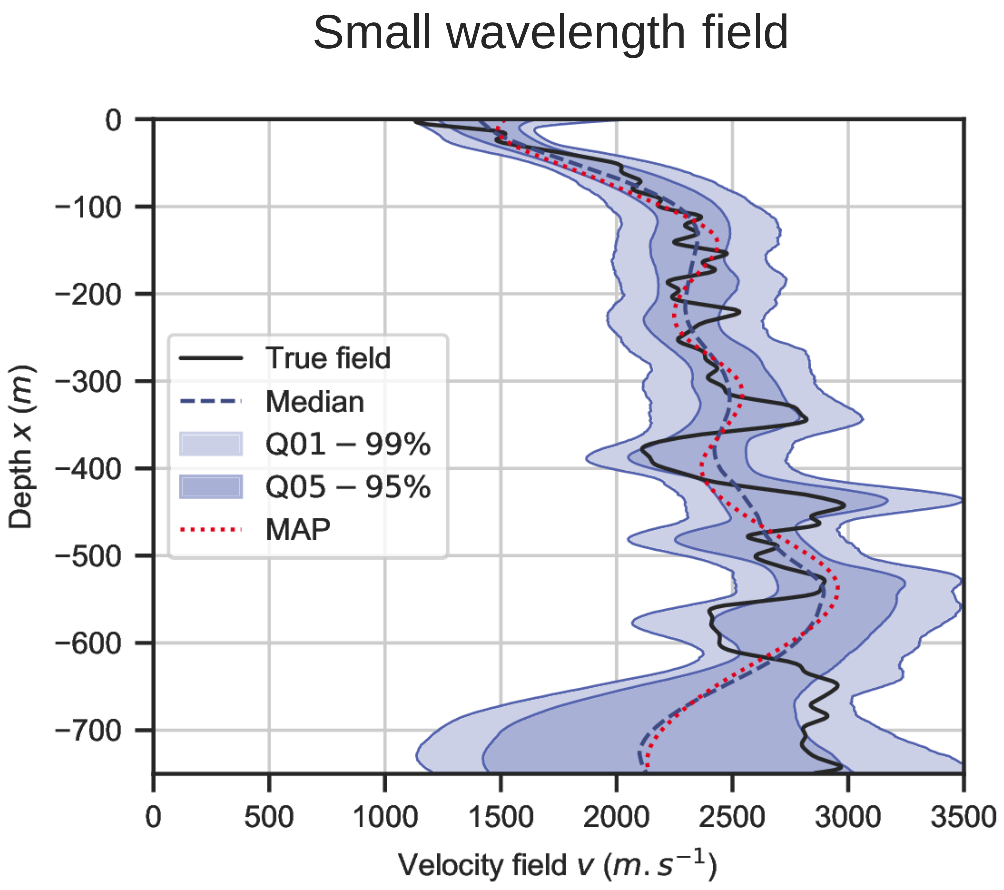
$l = 10$

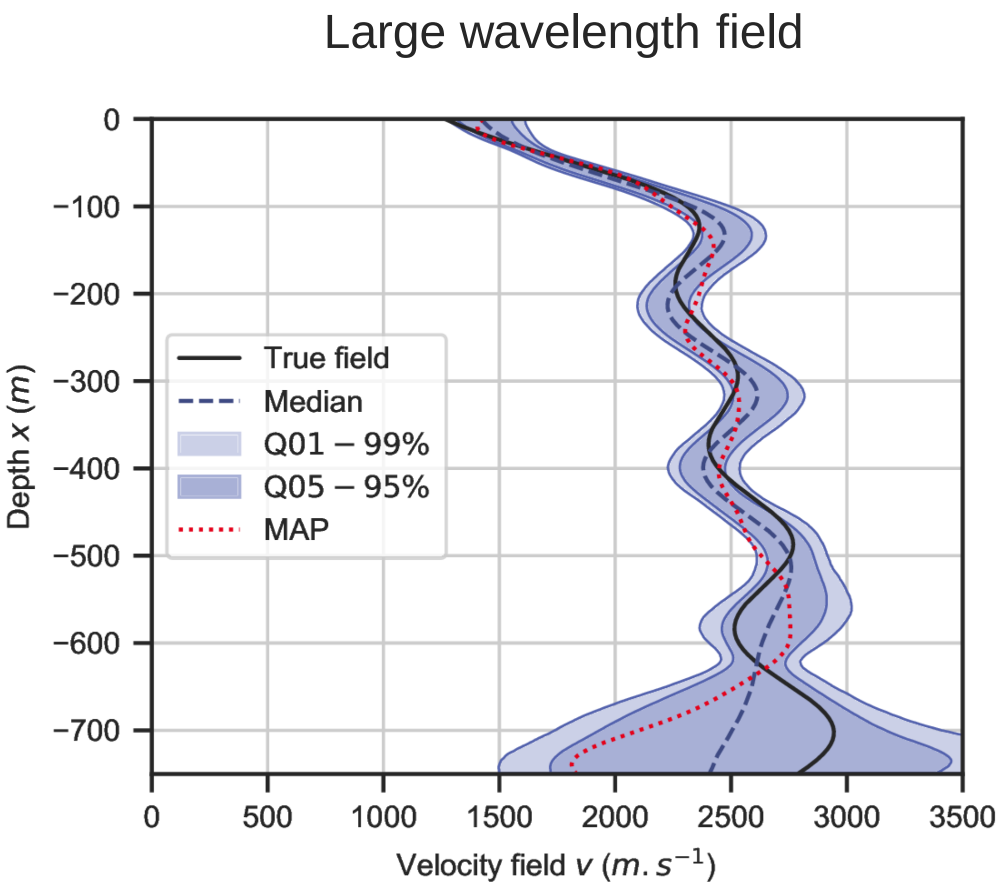
$l = 80$



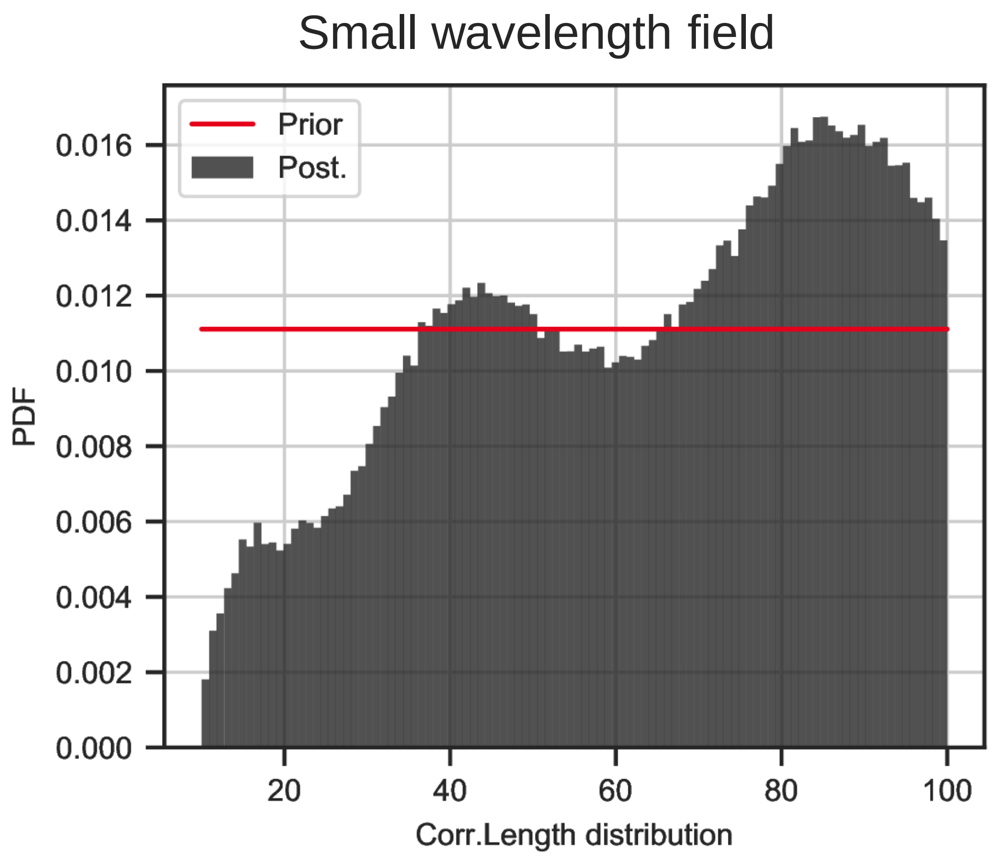
⇒ Using the same basis for both fields does not allow to distinguish them

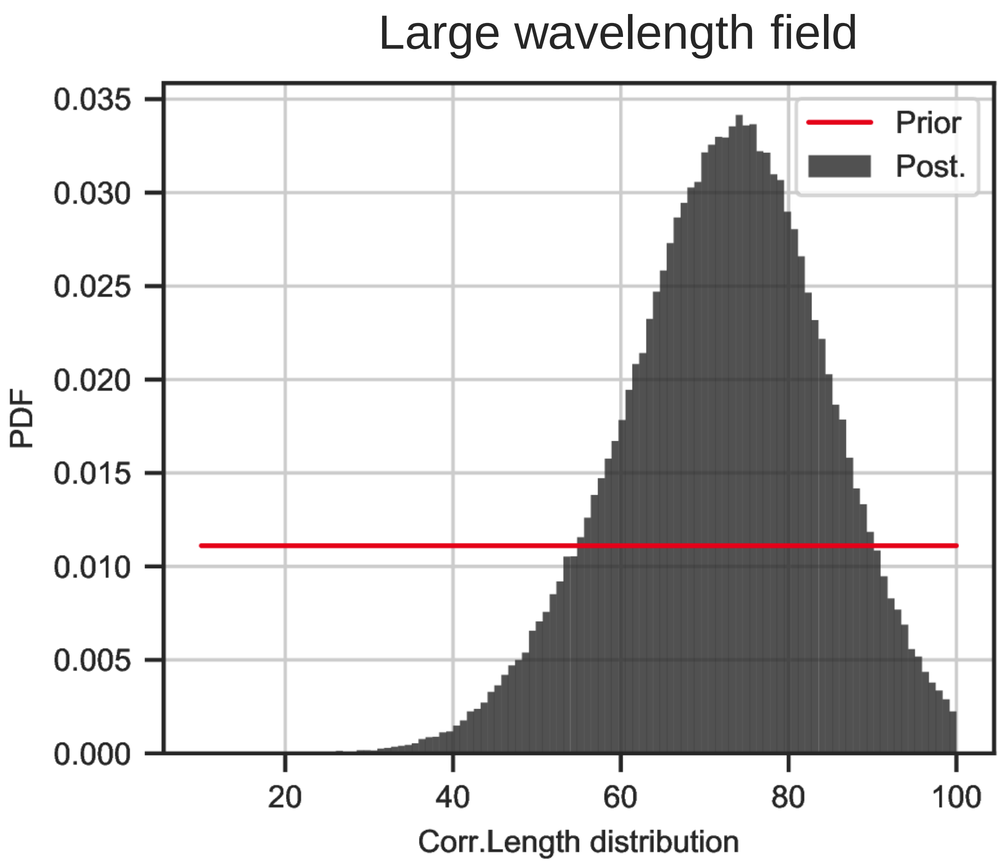
4. Results: with the change of measure method





4. Results: with the change of measure method





⇒ The coupled inference allows testing various field shapes but is not intended to select a 'best' hyperparameter value

Conclusion

- Change of measure: efficient algorithm for velocity field inference
  - Dimension reduction
  - Enlarge a priori parametrization: uncertainties are less ruled by the model selection
  - Without large computational cost increase
- Generalizable to other inverse problems
- Uncertainty propagation to other quantities (eg location)
- Development of adaptive methods

Thank you !

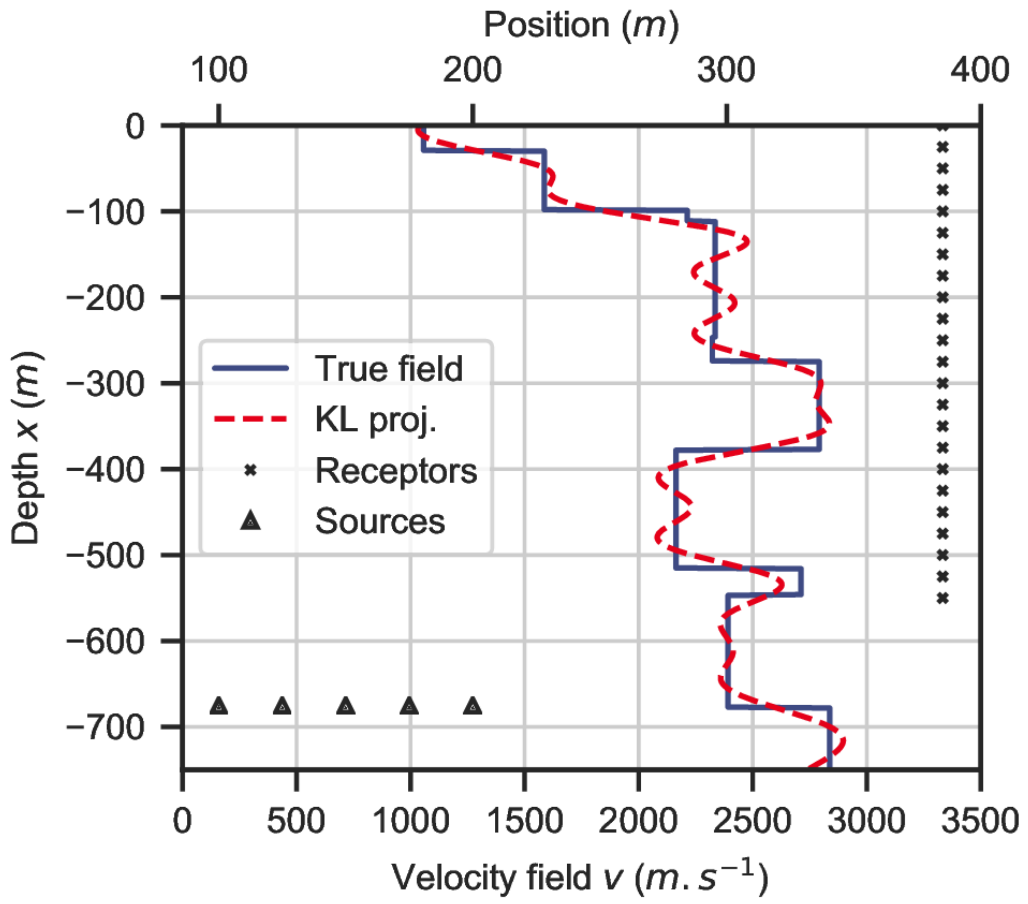
nadege.polette@cea.fr

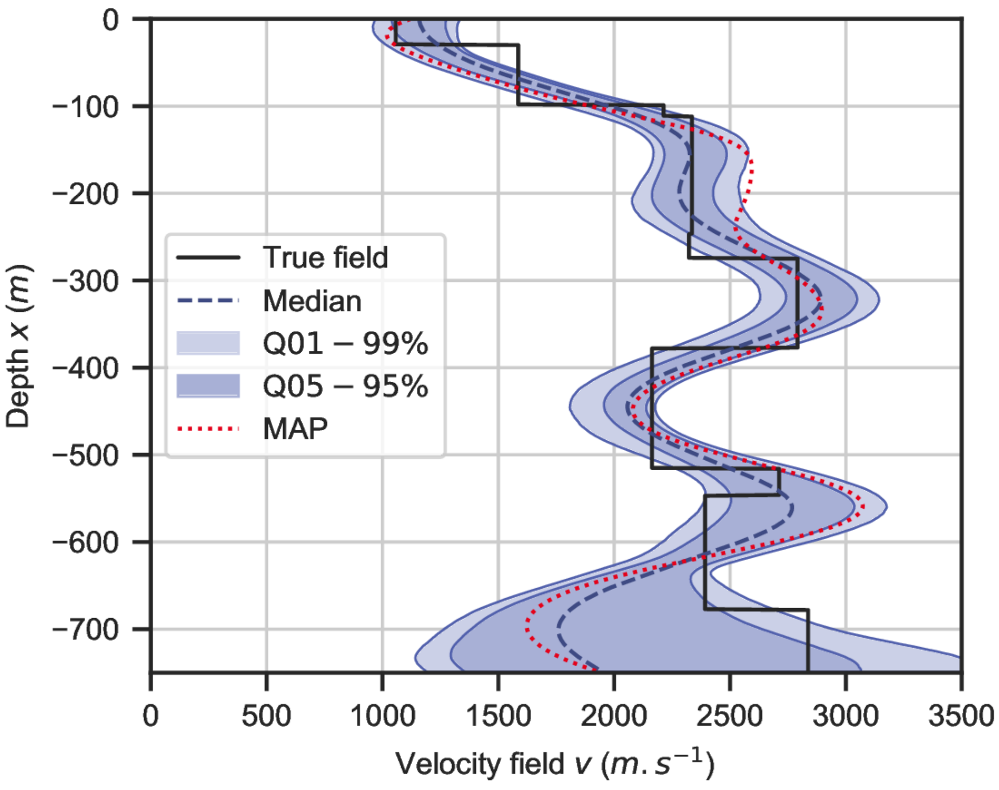
Keywords: inverse problem, (hierarchical) Bayesian inference, surrogate models (polynomial chaos), Markov Chain Monte Carlo, Dimension reduction, Karhunen-Loève decomposition

Bibliography

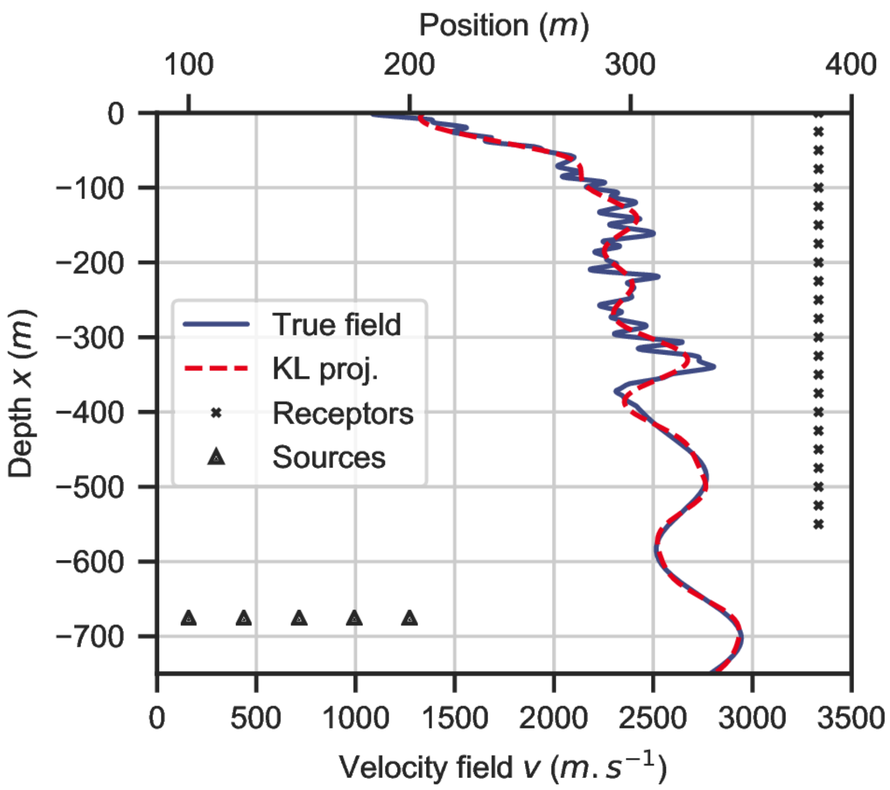
[Polette et al, 2024] N. Polette, O. Le Maître, P. Sochala, A. Gesret.
Change of Measure for Bayesian Field Inversion with Hierarchical Hyperparameters Sampling. *Preprint*, 2024
[Tarantola, 2005] A. Tarantola, Inverse Problem Theory and Methods for Model Parameter Estimation, Vol. xii, SIAM, 2005.
[Sivia, Skilling, 2006] D. Sivia, J. Skilling, Data Analysis: A Bayesian Tutorial, Oxford science publications, 2006.
[Doucet et al., 2013] A. Doucet, A. Smith, N. de Freitas, N. Gordon, Sequential Monte Carlo Methods in Practice, Information Science and Statistics, Springer NY, 2013.
[Wiener, 1938] N. Wiener, The Homogeneous Chaos, Am. J. Math, 1938.
[Ghanem, Spanos, 1991] R. G. Ghanem, P. D. Spanos, Stochastic Finite Element Method: Response Statistics, Springer NY, 1991.
[Xiu, Karniadakis, 2002] D. Xiu, G. E. Karniadakis, The Wiener Askey Polynomial Chaos for Stochastic Dierential Equations, SIAM JSC, 2002.
[Bodin et al., 2012] T. Bodin, M. Sambridge, N. Rawlinson, P. Arroucau, Transdimensional tomography with unknown data noise, GJI, 2012.
[Piana Agostinetti et al., 2015] N. Piana Agostinetti, G. Giacomuzzi, A. Malinverno, Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, GJI, 2015.
[Belhadj et al., 2018] J. Belhadj, T. Romary, A. Gesret, M. Noble, B. Figliuzzi, New parameterizations for Bayesian seismic tomography, Inverse Problems, 2018.
[Marzouk, Najm, 2009] Y. Marzouk, H. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, JCP, 2009.
[Karhunen, 1946] K. Karhunen, Zur spektraltheorie stochastischer prozesse, Ann. Acad. Sci. Fenn., 1946.
[Loeve, 1977] M. Loève, Probability Theory I, Vol. 45 of Graduate Texts in Mathematics, Springer NY, 1977.
[Tagade, Choi, 2014] P. M. Tagade, H.-L. Choi, A Generalized Polynomial Chaos-Based Method for Efficient Bayesian Calibration of Uncertain Computational Models, Inverse Problems in Science and Engineering, 2014.
[Sraj et al., 2016] I. Sraj, O. P. Le Maître, O. M. Knio, I. Hoteit, Coordinate transformation and Polynomial Chaos for the Bayesian inference of a Gaussian process with parametrized prior covariance function, CMAME, 2016
[Rasmussen, Williams, 2005] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005.
[Betancourt, Girolami, 2013] M. J. Betancourt, M. Girolami, Hamiltonian Monte Carlo for Hierarchical Models, 2013.

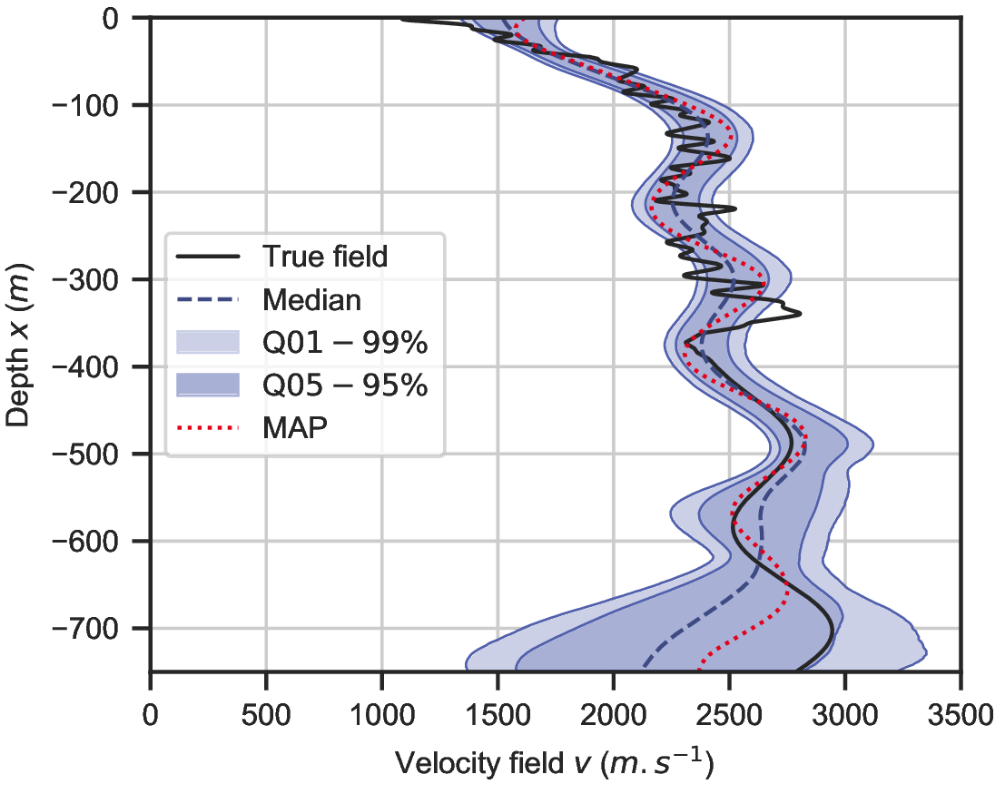
Appendix : discrete field





Appendix : non stationnary field





Appendix : non stationnary field