Field uncertainties estimation through (hyper)parameters sampling using Bayesian inference

PhD student : Nadège Polette[1,2],
Supervisors : Pierre Sochala[1], Alexandrine Gesret[2], Olivier Le Maître[3]

[1] CEA, DAM, DIF, F-91297 Arpajon, France
[2] Mines Paris PSL, centre de Géosciences, Fontainebleau, France
[3] CMAP, CNRS, Inria, École Polytechnique, IPP, Palaiseau, France

ETICS2023 - Cap Féret

Context

Detection and analysis of seismic events

Global scale

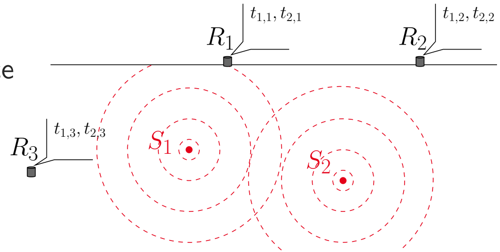
- International treaties (CTBT, NTP)
- Environment monitoring (IMS)

Regional scale

- Tsunami and seism alerts
- Risk prevention

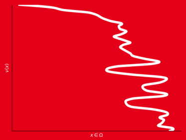
Local scale

- Knowledge of subsurface
- Exploitation



Context: seismic tomography

Quantity of interest



Velocity field f

Forward problem →

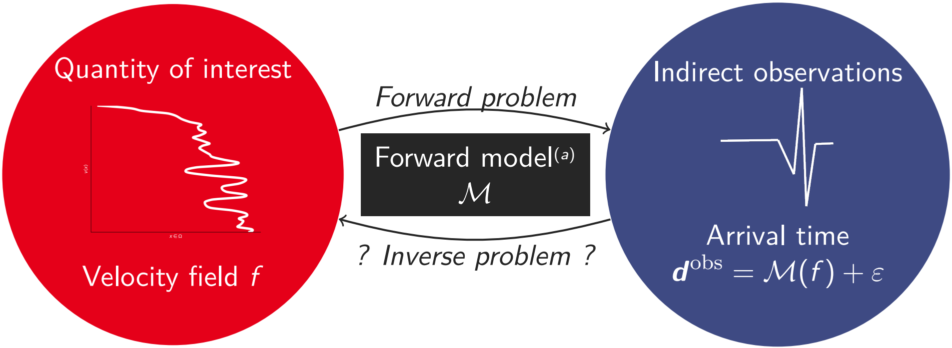
Forward model[a]
$\mathcal{M}$

Indirect observations

Arrival time
$\boldsymbol{d}^{\text{obs}} = \mathcal{M}(f) + \varepsilon$

[a] Eikonal solver [**noble2011**, **noble2011**]

Context: seismic tomography



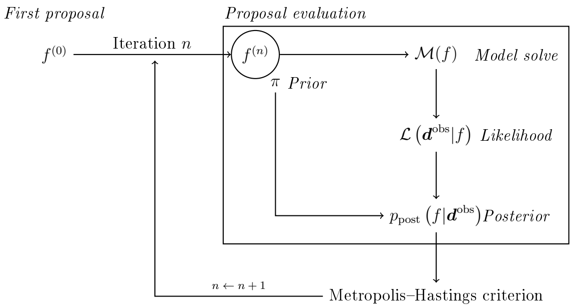
(a) Eikonal solver [noble2011, noble2011]

Objective: Estimation of a field (i) accurate,
(ii) with uncertainties,
(iii) fast

Bayes formulation

Bayes rule: $$p_{\text{post}}(f|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|f)\pi_{\mathcal{F}}(f)$$

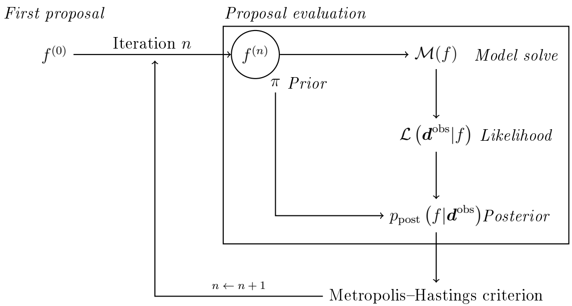
Markov Chain Monte–Carlo algorithm:



Bayes formulation

Bayes rule: $p_{\text{post}}(f|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|f)\pi_{\mathcal{F}}(f)$

Markov Chain Monte–Carlo algorithm:



$\Rightarrow$ Representation of f ?

Evaluation of $\mathcal{M}$?

$\rightarrow$ Polynomial chaos surrogate [**marzouk2009**, **marzouk2009**]

Field representation

$f(\boldsymbol{x})$ is seen as a particular *realization of a random (Gaussian) process* $\mathcal{G} \sim \mathcal{N}(0, k)$, where k is the *autocovariance function*

Karhunen–Loève decomposition

$$f(\boldsymbol{x}) = \mathcal{G}(\boldsymbol{x}, \theta) \simeq \sum_{i=1}^{r} \lambda_i^{1/2} u_i(\boldsymbol{x}) \eta_i(\theta), \text{ with } \eta_i = \lambda_i^{-1/2} \langle u_i, \mathcal{G} \rangle_\Omega$$

- $(u_i, \lambda_i)_{i \in \mathbb{N}^*}$ eigenelements of k:

$$\langle k(\boldsymbol{x}, \cdot), u_i \rangle_\Omega := \int_\Omega k(\boldsymbol{x}, \boldsymbol{x}') u_i(\boldsymbol{x}') d\boldsymbol{x}' = \lambda_i u_i(\boldsymbol{x})$$

- *Bi-orthonormality* of the decomposition:
  - $\forall i, j \in \mathbb{N}^*$, u_i, u_j orthonormal, $\langle u_i, u_j \rangle_\Omega = \delta_{i,j}$,
  - $\boldsymbol{\eta} := (\eta_i)_{1 \leqslant i \leqslant r} \sim \mathcal{N}(0, \mathrm{I}_r)$ (if Gaussian process)

$$\Rightarrow p_{\text{post}}(\boldsymbol{\eta} | \boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}} | \boldsymbol{\eta}) \pi(\boldsymbol{\eta})$$

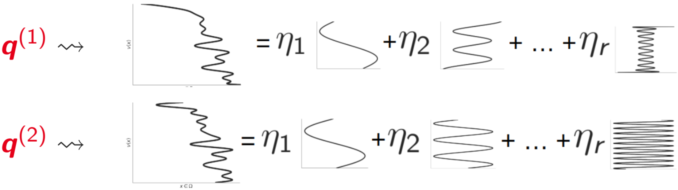
[**karhunen1946**, **karhunen1946**][**loeve1977**, **loeve1977**]

Dependency on hyperparameters

In fact, $\mathcal{G} \sim \mathcal{N}(0, k(\boldsymbol{q}))$,

$$f(\boldsymbol{x}) \simeq \sum_{i=1}^{r} \lambda_i^{1/2}(\boldsymbol{q}) u_i(\boldsymbol{x}, \boldsymbol{q}) \eta_i, \text{ with } \eta_i = \lambda_i^{-1/2}(\boldsymbol{q}) \langle u_i(\cdot, \boldsymbol{q}), f \rangle_\Omega$$

Squared exponential kernel: $k(x, y, \boldsymbol{q} := \{A, \ell\}) = A \exp\left(-\frac{\|x-y\|^2}{2\ell^2}\right)$



$$\Rightarrow p_{\mathrm{post}}(\boldsymbol{\eta}, \boldsymbol{q} | \boldsymbol{d}^{\mathrm{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\mathrm{obs}} | \boldsymbol{\eta}, \boldsymbol{q}) \pi(\boldsymbol{\eta}, \boldsymbol{q})$$

Change of measure method

Change of measure:

- Reference kernel $\overline{k}$ and associated basis $(\overline{\lambda}_i, \overline{u}_i)_{i\in\mathbb{N}^*}$ [**sraj2016**, **sraj2016**]
- Sample $(\boldsymbol{\xi}, \boldsymbol{q})$: *the* ***q****-dependency is transferred to the coordinates law,* [NP/Sochala/Le Maître/Gesret, in prep.]

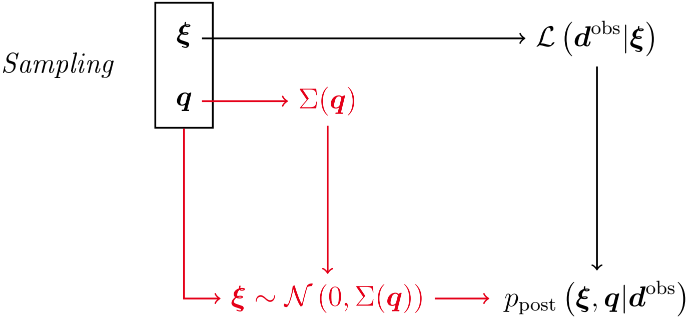
$$f(x) \simeq \overline{\mathcal{G}}^r(\boldsymbol{x}, \theta) := \sum_{i=1}^{r} \overline{\lambda}_i^{1/2} \overline{u}_i(\boldsymbol{x}) \xi_i(\theta) \text{ with } \boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$$

$$\text{and } p_{\text{post}}(\boldsymbol{\xi}, \boldsymbol{q} | \boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}} | \boldsymbol{\xi}) \pi(\boldsymbol{\xi} | \boldsymbol{q}) \pi(\boldsymbol{q})$$

The covariance matrix $\Sigma(\boldsymbol{q})$ writes

$$\forall 1 \leqslant i, j \leqslant r,\ \forall \boldsymbol{q} \in \mathbb{H}, \quad \Sigma_{ij}(\boldsymbol{q}) := (\overline{\lambda}_i \overline{\lambda}_j)^{-1/2} \left\langle \left\langle k(\cdot, \cdot, \boldsymbol{q}), \overline{u}_j \right\rangle_\Omega, \overline{u}_i \right\rangle_\Omega$$

Workflow

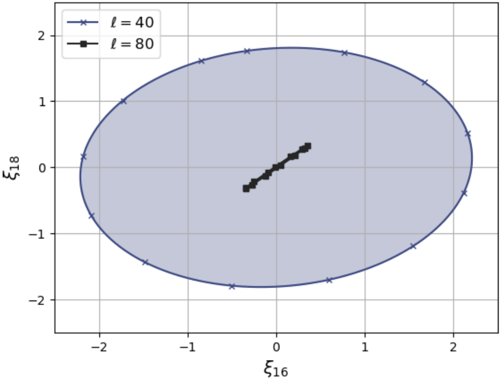


- $\Sigma_{ij}(\boldsymbol{q}) = (\overline{\lambda}_i\overline{\lambda}_j)^{-1/2}\left\langle\langle k(\cdot,\cdot,\boldsymbol{q}), \overline{u}_j\rangle_\Omega, \overline{u}_i\right\rangle_\Omega$

Sampling problem

Hierarchical formulation: $p_{\text{post}}(\xi, \boldsymbol{q}|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\xi)\pi(\xi|\boldsymbol{q})\pi(\boldsymbol{q})$,

$\xi \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$: *the prior distribution of ξ is highly sensitive to $\boldsymbol{q}$*



Projection of ξ priors for two different $\boldsymbol{q}$

Sampling strategy

Introduction of an *auxiliary variable* $\overline{\boldsymbol{\xi}}$ whose prior law does not depend on the hyperparameters (*e.g.* $\overline{\boldsymbol{\xi}} \sim \mathcal{N}(0, \Sigma_{\overline{\boldsymbol{\xi}}})$)

Algorithm 1 Sampling step for CoM method.

Input: $Y^{(n)} = (\overline{\boldsymbol{\xi}}^{(n)}, \boldsymbol{\xi}^{(n)}, \boldsymbol{q}^{(n)})$ ▷ Current state

1: Propose $(\overline{\boldsymbol{\xi}}^{*}, \boldsymbol{q}^{*}) \sim \mathcal{N}((\overline{\boldsymbol{\xi}}^{(n)}, \boldsymbol{q}^{(n)}), \widehat{C}^{(n)})$ ▷ Adaptive random walk

2: Set $\boldsymbol{\xi}^{*} = \Sigma(\boldsymbol{q}^{(n)})^{1/2} \Sigma_{\overline{\boldsymbol{\xi}}}^{-1/2} \overline{\boldsymbol{\xi}}^{*}$ ▷ Change of variable

3: Accept/Reject according to MH criterion on $p_{\text{post}}(\boldsymbol{\xi}^{*}, \boldsymbol{q}^{*})$.

- the ratio of the transition probabilities becomes

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

Surrogate quantities

Evaluating forward model $\mathcal{M}$ and derived quantities of $\Sigma(\boldsymbol{q})$ at each step is expensive $\Rightarrow$ use of *Polynomial Chaos (PC) expansions* [**wiener1938**, **wiener1938**][**ghanem1991**, **ghanem1991**][**xiu2002**, **xiu2002**]

$$m(\zeta) \simeq \widetilde{m}(\zeta) = \sum_{a \in \mathcal{A}} m_a \Psi_a(\zeta),$$

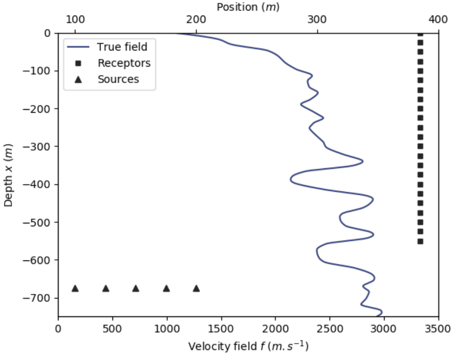
- Forward model: $\mathcal{M} \circ f(\boldsymbol{\xi}) = \widetilde{\mathcal{M}}(\boldsymbol{\xi})$
- $\Sigma(\boldsymbol{q})^{-1} = \widetilde{\Sigma^{-1/2}}(\boldsymbol{q})\widetilde{\Sigma^{-1/2}}(\boldsymbol{q})$
- $\log\det[\Sigma(\boldsymbol{q})] = (\widetilde{\log\det\Sigma})(\boldsymbol{q})$
- $\Sigma(\boldsymbol{q})^{1/2} = \widetilde{\Sigma^{1/2}}(\boldsymbol{q})$

$\Sigma^{\pm 1/2}$ is obtained by eigendecomposition of Σ

$$\Sigma^{\pm 1/2} = U\Lambda^{\pm 1/2}U^{\top},$$

where (U, Λ) are the eigenelements of Σ

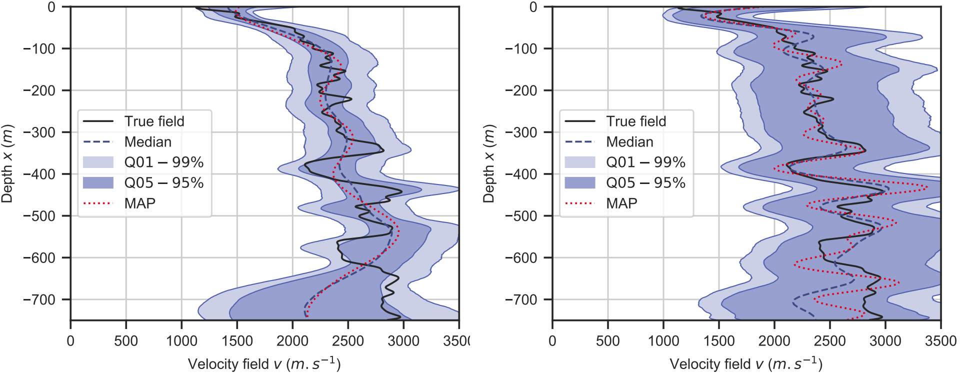
Application to seismic tomography



Application case: 1D section of Amoco model [**obrien1994**, **obrien1994**] and location of stations

$\boldsymbol{d}^{\mathrm{obs}}$: time of arrival, with noise level $\alpha = 0.002s$
$r = 20$, $\boldsymbol{q} = \{A, \ell\}$

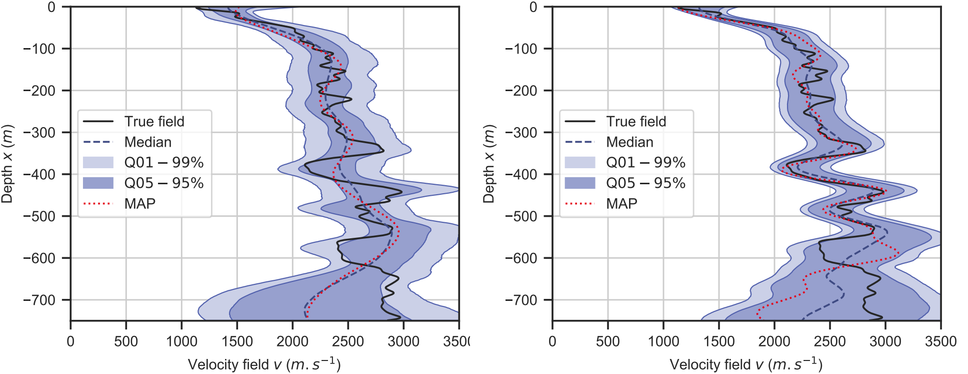
Inference of a field (SW)



(a) Proposed method (b) $\ell = 10$

Comparison of inference results for different bases - small wavelength field

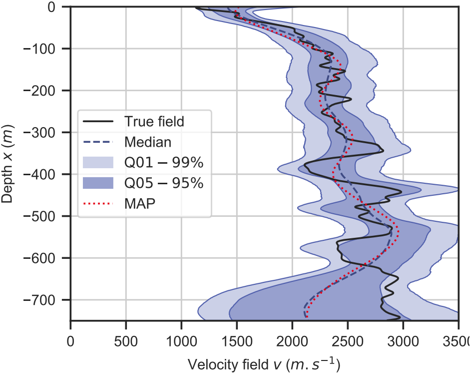
Inference of a field (SW)



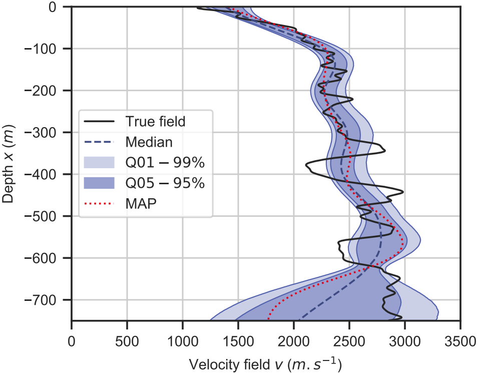
(a) Proposed method (b) $\ell = 34$

Comparison of inference results for different bases - small wavelength field

Inference of a field (SW)



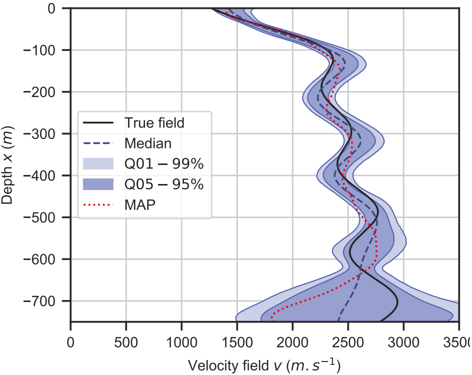
(a) Proposed method



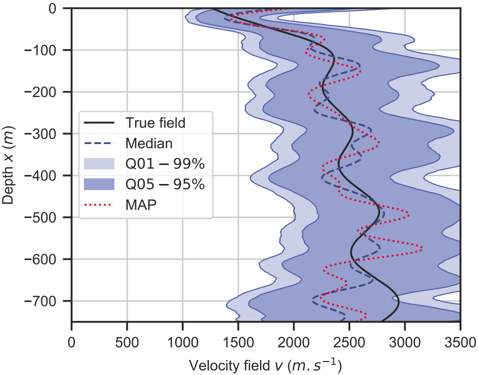
(b) $\ell = 80$

Comparison of inference results for different bases - small wavelength field

Inference of a field (LW)



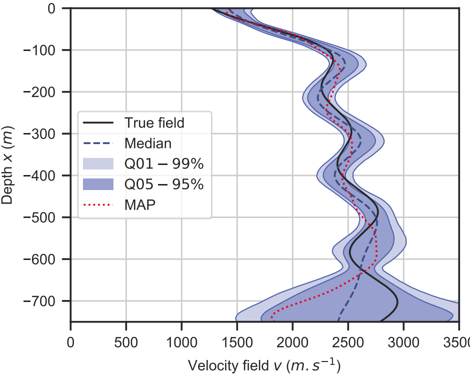
(a) Proposed method



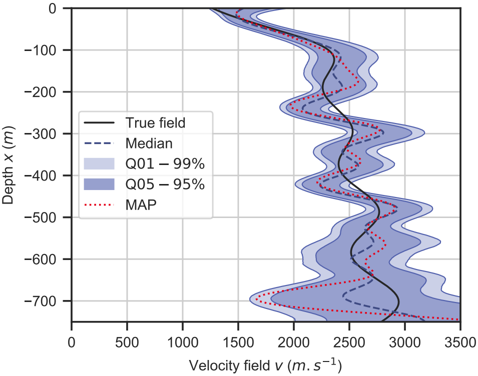
(b) $\ell = 10$

Comparison of inference results for different bases - large wavelength field

Inference of a field (LW)



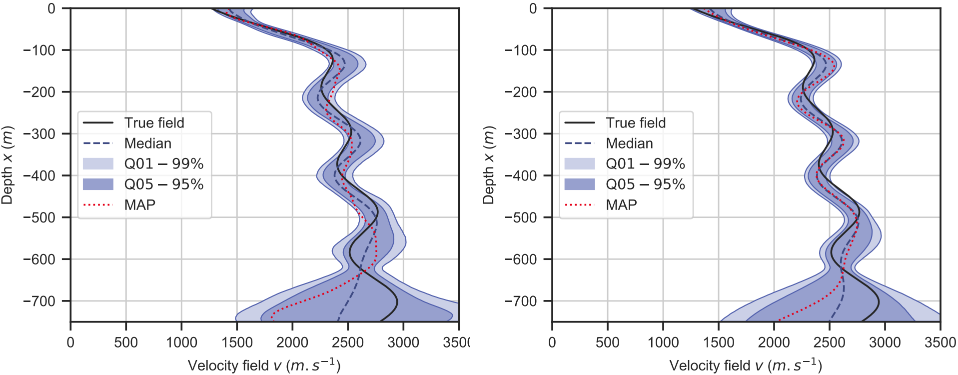
(a) Proposed method



(b) $\ell = 34$

Comparison of inference results for different bases - large wavelength field

Inference of a field (LW)

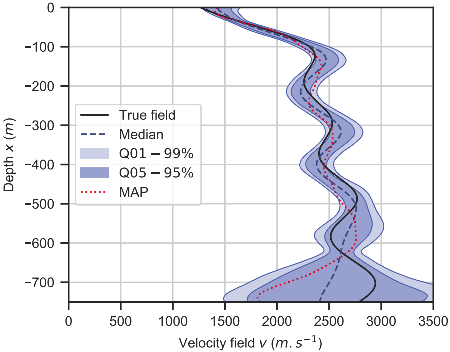


(a) Proposed method

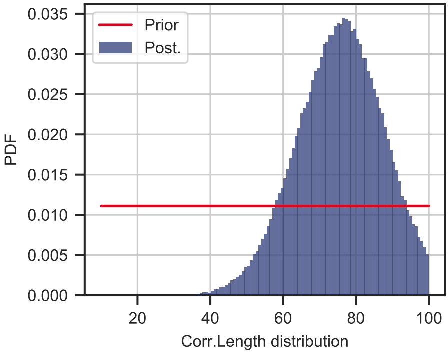
(b) $\ell = 80$

Comparison of inference results for different bases - large wavelength field

Hyperparameters distribution (LW)



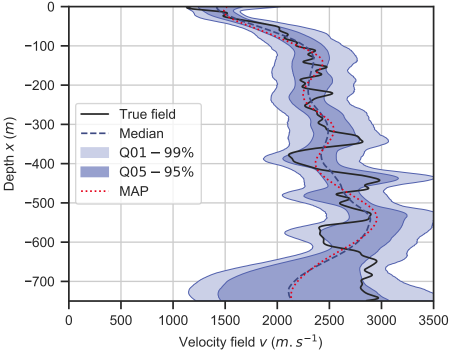
(a) Field posterior distribution



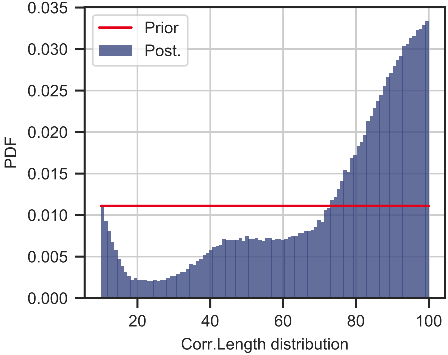
(b) Correlation length posterior distribution

Posterior distributions - large wavelength field

Hyperparameters distribution (SW)



(a) Field posterior distribution

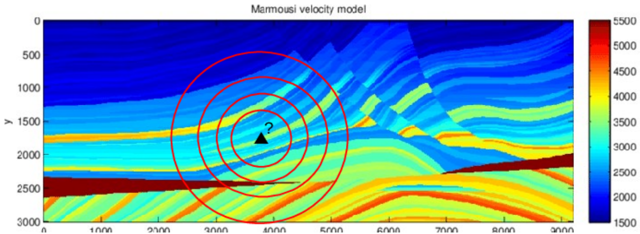


(b) Correlation length posterior distribution

Posterior distributions - small wavelength field

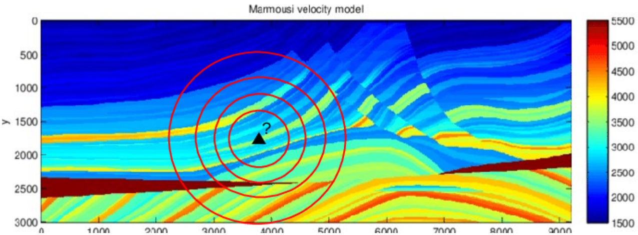
Conclusion and perspectives

- Change of measure
  - Dimension reduction of the field
  - Enlarge *a priori* parametrization $\rightsquigarrow$ uncertainties are less ruled by the model selection
  - Without large computational cost increase



Conclusion and perspectives

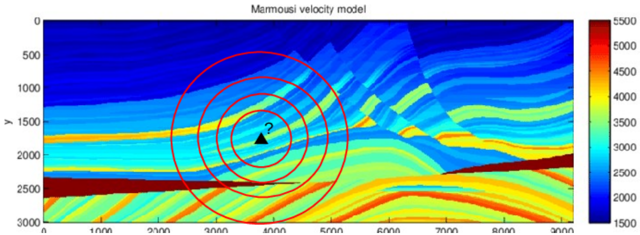
- Change of measure
  - Dimension reduction of the field
  - Enlarge *a priori* parametrization $\rightsquigarrow$ uncertainties are less ruled by the model selection
  - Without large computational cost increase
- Development of adaptive methods
- Use of the posterior quantity to propagate uncertainty to other parameters (e.g. seismic location)
- *How to quantify the relevance of the uncertainties ?*



Conclusion and perspectives

- Change of measure
  - Dimension reduction of the field
  - Enlarge *a priori* parametrization $\rightsquigarrow$ uncertainties are less ruled by the model selection
  - Without large computational cost increase
- Development of adaptive methods
- Use of the posterior quantity to propagate uncertainty to other parameters (e.g. seismic location)
- *How to quantify the relevance of the uncertainties ?*

Thank you !



References I

Polynomial Chaos expansions

Evaluating forward model and derived quantities of $\Sigma(\boldsymbol{q})$ at each step is expensive $\Rightarrow$ use of *Polynomial Chaos (PC) expansions* [**wiener1938**, **wiener1938**][**ghanem1991**, **ghanem1991**][**xiu2002**, **xiu2002**]

$$m(\zeta) \simeq \widetilde{m}(\zeta) = \sum_{a \in \mathcal{A}} m_a \Psi_a(\zeta),$$

- $\{\Psi_a(\zeta)\}_{a \in \mathcal{A}}$: orthogonal polynomials in $L^2(\zeta)$
- $\{m_a\}_{a \in \mathcal{A}}$: PC coefficients
- $\mathcal{A}$: set of multi-indexes

Building framework:

- Exact evaluations at quadrature points (Gauss–Legendre or sparse grids)
- Computations of PC coefficients by matrix-vector product

Surrogate quantities

Evaluating forward model and derived quantities of $\Sigma(\boldsymbol{q})$ at each step is expensive $\Rightarrow$ use of *Polynomial Chaos (PC) expansions* [**wiener1938**, **wiener1938**][**ghanem1991**, **ghanem1991**][**xiu2002**, **xiu2002**]

$$m(\zeta) \simeq \widetilde{m}(\zeta) = \sum_{a \in \mathcal{A}} m_a \Psi_a(\zeta),$$

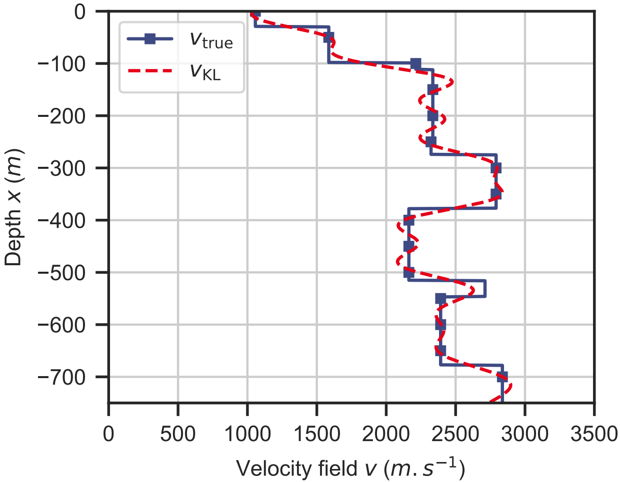
- Forward model: $\mathcal{M} \circ f(\boldsymbol{\xi}) = \widetilde{\mathcal{M}}(\boldsymbol{\xi})$
- $\Sigma(\boldsymbol{q})^{-1} = \widetilde{\Sigma^{-1/2}}(\boldsymbol{q})\widetilde{\Sigma^{-1/2}}(\boldsymbol{q})$
- $\log\det\left[\Sigma(\boldsymbol{q})\right] = \left(\widetilde{\log\det\Sigma}\right)(\boldsymbol{q})$
- $\Sigma(\boldsymbol{q})^{1/2} = \widetilde{\Sigma^{1/2}}(\boldsymbol{q})$

$\Sigma^{\pm 1/2}$ is obtained by eigendecomposition of Σ

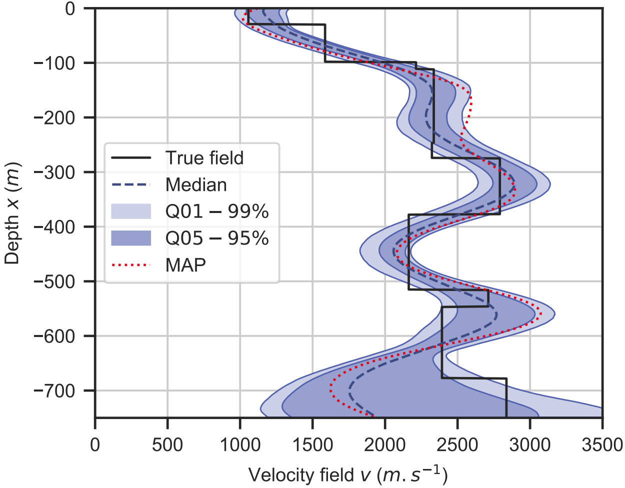
$$\Sigma^{\pm 1/2} = U \Lambda^{\pm 1/2} U^\top,$$

where (U, Λ) are the eigenelements of Σ

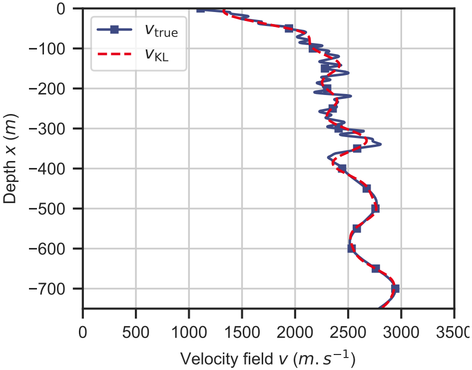
Projection of a discrete field

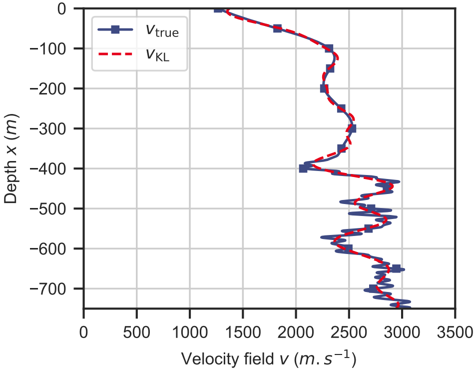


Inference of a discrete field

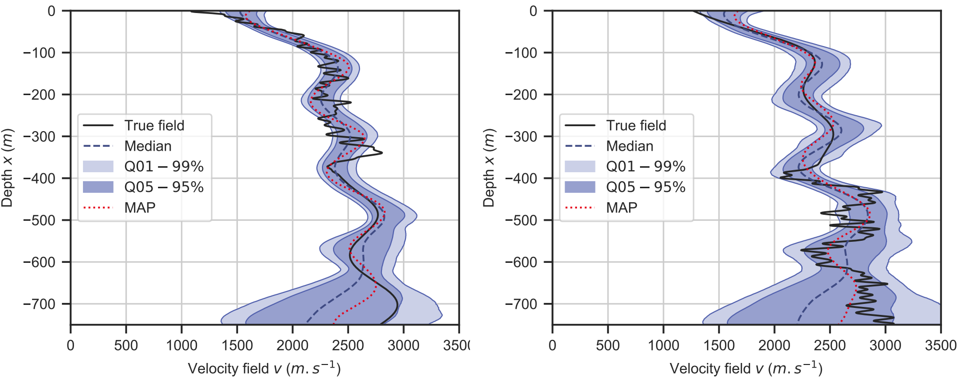


Projection of a non-stationnary field



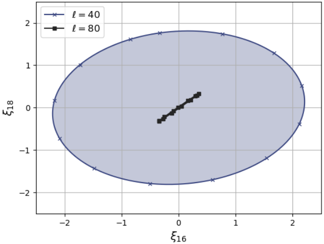


Inference of a non-stationnary field



Sampling problem

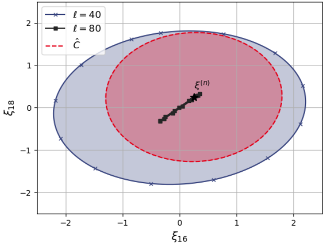
Hierarchical formulation: $p_{\text{post}}(\boldsymbol{\xi}, \boldsymbol{q}|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{q})\pi(\boldsymbol{q})$,
$\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$: *the prior distribution of $\boldsymbol{\xi}$ is highly sensitive to $\boldsymbol{q}$*



Projection of $\boldsymbol{\xi}$ priors for two different $\boldsymbol{q}$

Sampling problem

Hierarchical formulation: $p_{\text{post}}(\boldsymbol{\xi}, \boldsymbol{q}|\boldsymbol{d}^{\text{obs}}) \propto \mathcal{L}(\boldsymbol{d}^{\text{obs}}|\boldsymbol{\xi})\pi(\boldsymbol{\xi}|\boldsymbol{q})\pi(\boldsymbol{q})$,
$\boldsymbol{\xi} \sim \mathcal{N}(0, \Sigma(\boldsymbol{q}))$: *the prior distribution of $\boldsymbol{\xi}$ is highly sensitive to $\boldsymbol{q}$*



Projection of $\boldsymbol{\xi}$ priors for two different $\boldsymbol{q}$

Initial sampling : Metropolis–Hastings (MH) random walk with adapted covariance proposal

$$Y^\star := (\boldsymbol{\xi}^\star, \boldsymbol{q}^\star) \sim \mathcal{N}(Y^{(n)}, \widehat{C}^{(n)}), \text{with } \widehat{C}^{(n)} \propto \mathbb{C}\text{ov}\left(Y^{(1)}, \dots, Y^{(n)}\right)$$