Automated data-driven inverse problem resolution: Applications in microfluidics and epidemiology

Barbara Pascal, bpascal-fr.github.io

Joint work with Patrice Abry, Nelly Pustelnik, Valérie Vidal, and Samuel Vaiter

January 30, 2026

AI WILD West, Rennes, France

Observation model

$$\boldsymbol{y} \sim \mathcal{B}\left(\boldsymbol{\Phi}\overline{\boldsymbol{x}}\right)$$

- $\boldsymbol{y} \in \mathbb{R}^P$: degraded observations;
- $\overline{\boldsymbol{x}} \in \mathbb{R}^N$: unknown quantity of interest;
- $\boldsymbol{\Phi} : \mathbb{R}^N \to \mathbb{R}^P$: known deformation;
- $\mathcal{B}$: random measurement noise.

Observation model

$$\boldsymbol{y} \sim \mathcal{B}(\boldsymbol{\Phi}\overline{\boldsymbol{x}})$$

- $\boldsymbol{y} \in \mathbb{R}^P$: degraded observations;
- $\overline{\boldsymbol{x}} \in \mathbb{R}^N$: unknown quantity of interest;
- $\boldsymbol{\Phi} : \mathbb{R}^N \to \mathbb{R}^P$: known deformation;
- $\mathcal{B}$: random measurement noise.

Goal: Estimate $\overline{\boldsymbol{x}}$

Observation model

$$\boldsymbol{y} \sim \mathcal{B}(\boldsymbol{\Phi}\overline{\boldsymbol{x}})$$

- $\boldsymbol{y} \in \mathbb{R}^P$: degraded observations;
- $\overline{\boldsymbol{x}} \in \mathbb{R}^N$: unknown quantity of interest;
- $\boldsymbol{\Phi} : \mathbb{R}^N \to \mathbb{R}^P$: known deformation;
- $\mathcal{B}$: random measurement noise.

Goal: Estimate $\overline{\boldsymbol{x}}$

- ill-conditioned, rank deficient $\boldsymbol{\Phi}$

Inpainting

(Guillemot et al., 2013, *IEEE Sig. Process. Mag.*)

Super-resolution

(Marquina et al., 2008, *J. Sci. Comput.*)

Deblurring

(Pan, 2016, *IEEE Trans. Pattern Anal. Mach. Intell.*)

Observation model

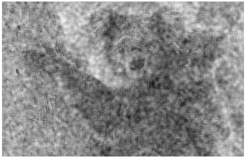
$$\boldsymbol{y} \sim \mathcal{B}(\boldsymbol{\Phi}\overline{\boldsymbol{x}})$$

- $\boldsymbol{y} \in \mathbb{R}^P$: degraded observations;
- $\overline{\boldsymbol{x}} \in \mathbb{R}^N$: unknown quantity of interest;
- $\boldsymbol{\Phi} : \mathbb{R}^N \to \mathbb{R}^P$: known deformation;
- $\mathcal{B}$: random measurement noise.

Goal: Estimate $\overline{\boldsymbol{x}}$

- ill-conditioned, rank deficient $\boldsymbol{\Phi}$
- correlated, data-dependent $\mathcal{B}$

Correlated



(Pascal et al., 2021, *J. Math. Imaging Vis.*)

Data-dependent

(Luisier et al., 2010, *IEEE Trans. Image Process.*)

Multiplicative

(Shama, 2016, *Appl. Math. Comput.*)

Variational estimator

$$\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \operatorname*{Argmin}_{\boldsymbol{x}\in\mathbb{R}^N} \mathcal{D}(\boldsymbol{y},\boldsymbol{\Phi}\boldsymbol{x})$$

- $\mathcal{D}(\boldsymbol{y};\cdot) = -\log \mathbb{P}(\boldsymbol{y}|\cdot)$: negative log-likelihood   Ex: $\mathcal{D}(\boldsymbol{y};\boldsymbol{x}) = \|\boldsymbol{y}-\boldsymbol{x}\|_2^2$

No regularization

$\mathcal{R} = 0$

Variational estimator

$$\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}}\, \mathcal{D}(\boldsymbol{y},\boldsymbol{\Phi x}) + \lambda\mathcal{R}(\boldsymbol{x})$$

- $\mathcal{D}(\boldsymbol{y};\cdot) = -\log\mathbb{P}(\boldsymbol{y}|\cdot)$: negative log-likelihood — Ex: $\mathcal{D}(\boldsymbol{y};\boldsymbol{x}) = \|\boldsymbol{y}-\boldsymbol{x}\|_2^2$
- $\mathcal{R}$: regularization term encoding a priori knowledge — Ex: $\mathcal{R}(\boldsymbol{x}) = \|\mathbf{D}_1\boldsymbol{x}\|_q^q$

(Giovannelli & Idier, 2015, *Wiley*)

No regularization	Smooth	Piecewise constant
		
$\mathcal{R} = 0$	$\mathcal{R}(\boldsymbol{x}) = \|\mathbf{D}_1\boldsymbol{x}\|_2^2$	$\mathcal{R}(\boldsymbol{x}) = \|\mathbf{D}_1\boldsymbol{x}\|_1$
	(Tikhonov et al., 1977, *Wiley*)	(Rudin et al., 1992, *Physica D*)

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \operatorname*{Argmin}_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{Argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

$\lambda = 25$

too regularized

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

$\boldsymbol{\lambda^\dagger = 5}$

optimal

$\lambda = 25$

too regularized

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\, \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

$\boldsymbol{\lambda^\dagger = 5}$

optimal

$\lambda = 25$

too regularized

Oracle-based hyperparameter selection

$$\lambda^\dagger \in \underset{\lambda\in\Lambda}{\mathrm{Argmin}}\, \mathcal{O}(\boldsymbol{y};\lambda)$$

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\, \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

$\boldsymbol{\lambda^\dagger = 5}$

optimal

$\lambda = 25$

too regularized

Oracle-based hyperparameter selection

$$\lambda^\dagger \in \underset{\lambda\in\Lambda}{\mathrm{Argmin}}\, \mathcal{O}(\boldsymbol{y};\lambda)$$

- golden case: $\mathcal{O}(\boldsymbol{y};\lambda) = \|\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) - \overline{\boldsymbol{x}}\|^2 \Longrightarrow$ efficient bi-level $(\boldsymbol{x},\lambda)$ minimization

Fine-tuning of the regularization parameter

Example: $\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{Argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \lambda\|\mathbf{D}_1\boldsymbol{x}\|_2^2$ (Tikhonov)

$\lambda = 1$

not enough

$\boldsymbol{\lambda^\dagger = 5}$

optimal

$\lambda = 25$

too regularized

Oracle-based hyperparameter selection

$$\lambda^\dagger \in \underset{\lambda\in\Lambda}{\operatorname{Argmin}}\, \mathcal{O}(\boldsymbol{y};\lambda)$$

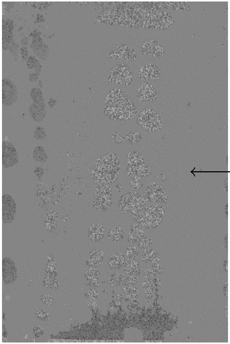
- golden case: $\mathcal{O}(\boldsymbol{y};\lambda) = \|\widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) - \overline{\boldsymbol{x}}\|^2 \Longrightarrow$ efficient bi-level $(\boldsymbol{x},\lambda)$ minimization
- practical case: ground truth $\overline{\boldsymbol{x}}$ **not available!** $\Longrightarrow$ data-driven $\mathcal{O}(\boldsymbol{y};\lambda)$

Image processing:

Texture segmentation

Multiphase flow through porous media

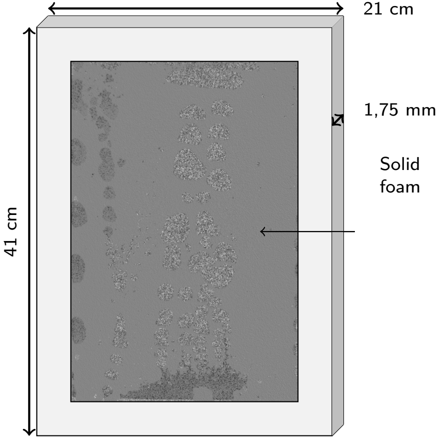
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



Solid
foam

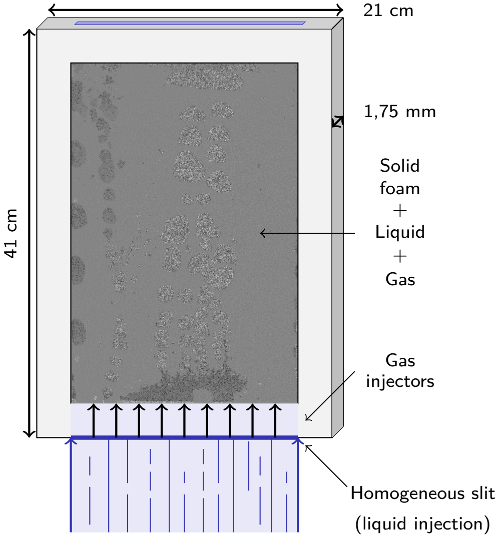
Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)

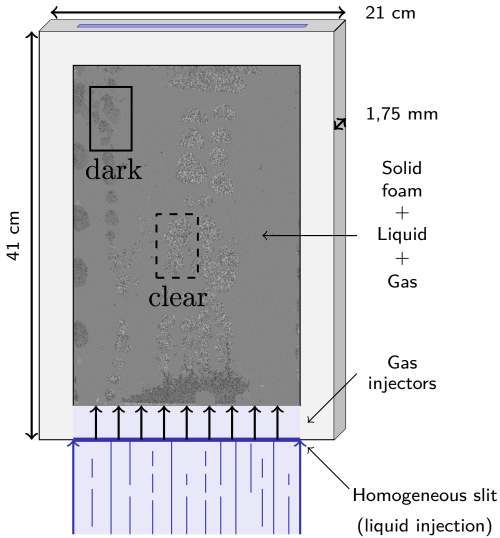


Multiphase flow through porous media

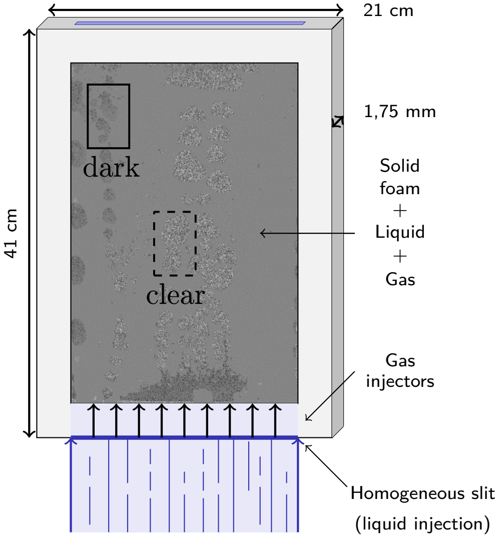
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



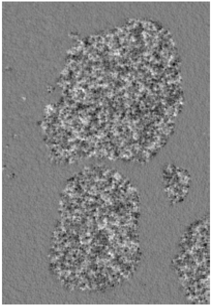
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



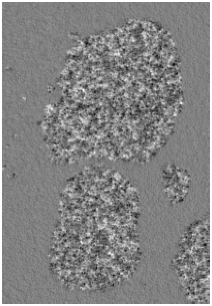
Multiphase flow through porous media
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)

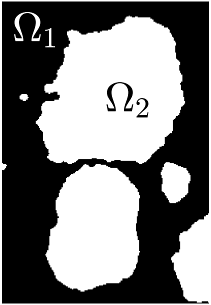


→ 1600 × 1100 pixels
→ video: ∼ 1000 images
→ phase diagram: ∼ 10 flow rates



Textured image segmentation

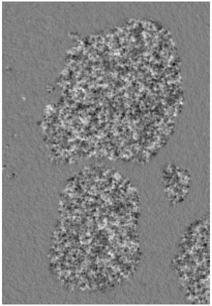


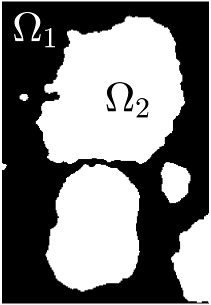


Goal: obtain a partition of the image into K homogeneous textures

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

Textured image segmentation

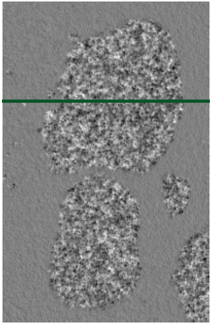


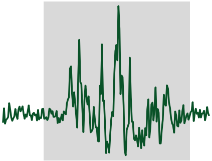


Goal: obtain a partition of the image into K **homogeneous textures**

$$\Omega = \Omega_1 \bigsqcup \ldots \bigsqcup \Omega_K$$

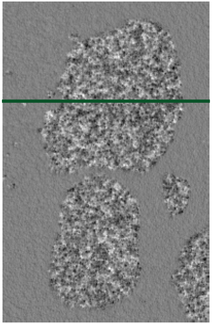
Piecewise monofractal model

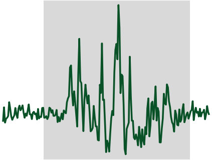




Fractal attributes

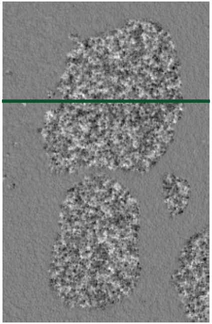
- variance σ^2 *amplitude of variations*

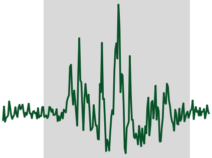




Fractal attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

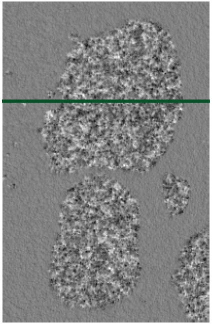


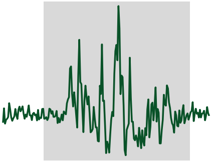


Fractal attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$



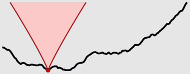


Piecewise monofractal model

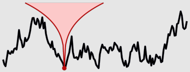
Fractal attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

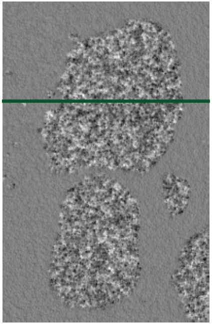
$$|f(x) - f(y)| \leq \sigma(x)|x - y|^{h(x)}$$

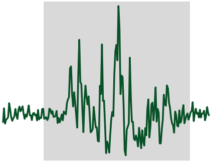


$h(x) \equiv h_1 = 0.9$



$h(x) \equiv h_2 = 0.3$





Fractal attributes

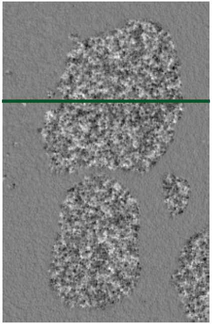
- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

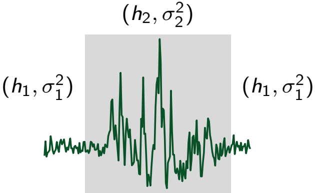
$$|f(x) - f(y)| \leq \sigma(x)|x-y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\qquad$ $h(x) \equiv h_2 = 0.3$

Segmentation

- σ^2 and h piecewise constant





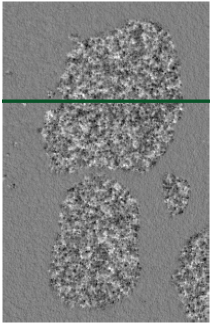
Piecewise monofractal model

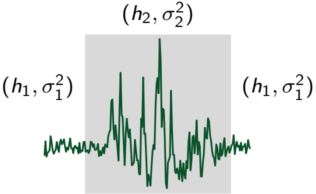
Fractal attributes

- variance σ^2 — *amplitude of variations*
- local regularity h — *scale invariance*

$$|f(x) - f(y)| \le \sigma(x)|x - y|^{h(x)}$$

$h(x) \equiv h_1 = 0.9$ $\qquad$ $h(x) \equiv h_2 = 0.3$

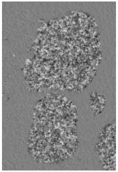




Segmentation

- σ^2 and h piecewise constant
- region Ω_k characterized by (σ_k^2, h_k)

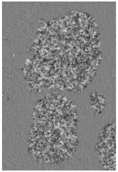
Textured image

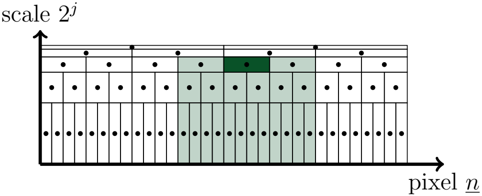


Multiscale analysis

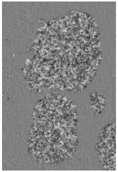
Textured image

Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$



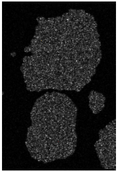


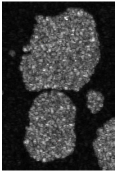
Textured image

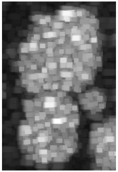


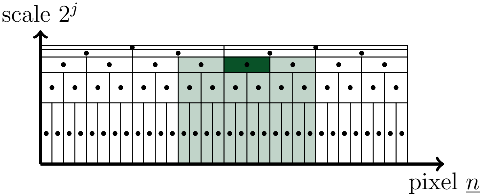
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



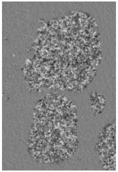




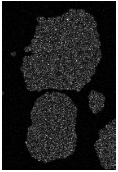


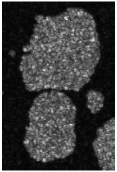
Textured image

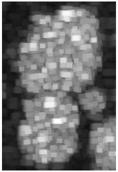
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$



Scale $a = 2^1$ $a = 2^2$ ... $a = 2^5$





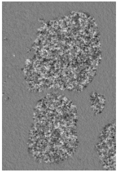


Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underset{\text{regularity}}{\boldsymbol{h}} + \underset{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}{\boldsymbol{v}}$$

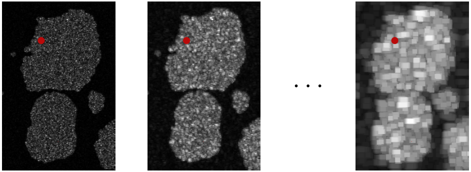
Multiscale analysis

Textured image



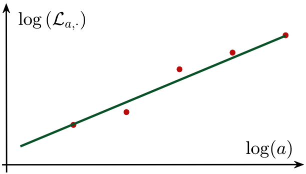
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



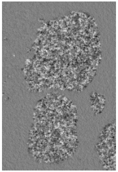
Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



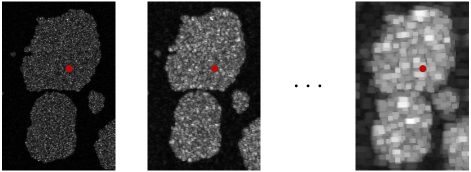
Multiscale analysis

Textured image



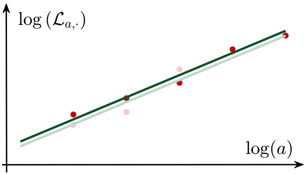
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

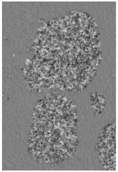
$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a \to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

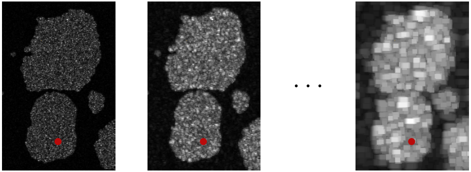


Textured image

Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

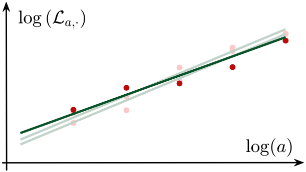
Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



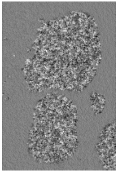


Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$

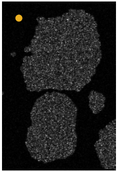


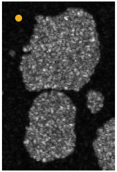
Textured image



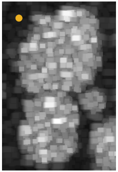
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



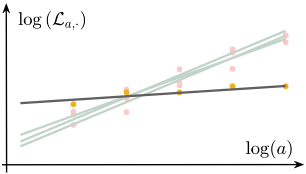


…



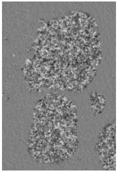
Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto\log(\boldsymbol{\sigma}^2)\\ \text{(variance)}}}$$



Multiscale analysis

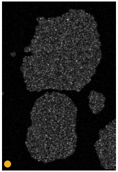
Textured image



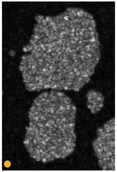
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale

$a = 2^1$

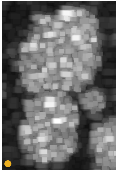


$a = 2^2$



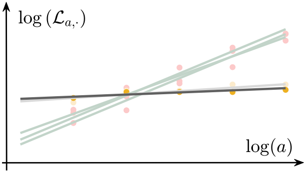
$\cdots$

$a = 2^5$



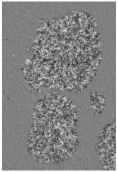
Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto \log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



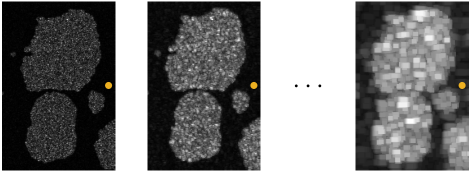
Multiscale analysis

Textured image



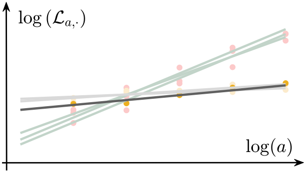
Local maximum of wavelet coefficients: $\mathcal{L}_{a,\cdot}$

Scale $a = 2^1$ $a = 2^2$ … $a = 2^5$



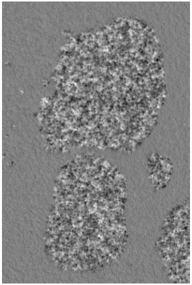
Proposition (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log\left(\mathcal{L}_{a,\cdot}\right) \underset{a\to 0}{\simeq} \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\substack{\propto\log(\boldsymbol{\sigma}^2) \\ \text{(variance)}}}$$



Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

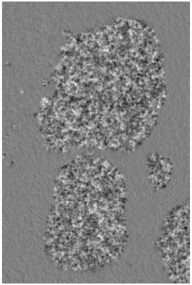
Textured image



Linear regression $\log\left(\mathcal{L}_{a,\cdot}\right) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\| \log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v} \right\|^2$$

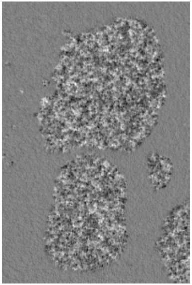
Textured image



Linear regression $\log(\mathcal{L}_{a,\cdot}) \simeq \log(a) \underbrace{\boldsymbol{h}}_{\text{regularity}} + \underbrace{\boldsymbol{v}}_{\propto \log(\boldsymbol{\sigma}^2)}$

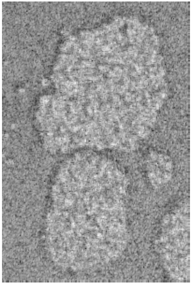
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\| \log(\mathcal{L}_{a,\cdot}) - \log(a)\boldsymbol{h} - \boldsymbol{v} \right\|^2$$

Textured image



Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$

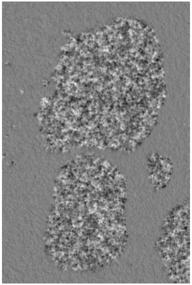


Direct punctual estimation

Linear regression $\underbrace{\mathbb{E}\log\left(\mathcal{L}_{a,\cdot}\right)}_{\text{expected value}} = \log(a)\underbrace{\bar{\boldsymbol{h}}}_{\text{regularity}} + \underbrace{\bar{\boldsymbol{v}}}_{\propto\log(\boldsymbol{\sigma}^2)}$

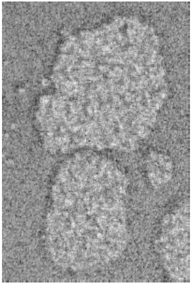
$$\left(\widehat{\boldsymbol{h}}^{\mathrm{LR}}, \widehat{\boldsymbol{v}}^{\mathrm{LR}}\right) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a=a_{\min}}^{a_{\max}} \left\|\log\left(\mathcal{L}_{a,\cdot}\right) - \log(a)\boldsymbol{h} - \boldsymbol{v}\right\|^2$$

Textured image



Local regularity $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

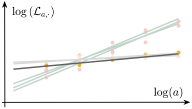
Local power $\widehat{\boldsymbol{v}}^{\mathrm{LR}}$



$\longrightarrow$ large estimation variance

$$\underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}}$$

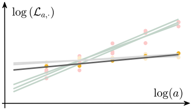
→ fidelity to the log-linear model

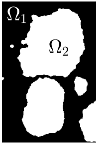


$$\underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model

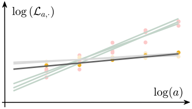
→ favors piecewise constancy



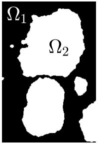


$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \;+\; \lambda \, \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

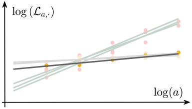
→ fidelity to the log-linear model

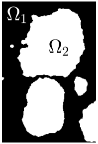


→ favors piecewise constancy



$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\substack{\textbf{Least-Squares} \\ \to \text{ fidelity to the log-linear model}}} + \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\substack{\textbf{Total Variation} \\ \to \text{ favors piecewise constancy}}}$$



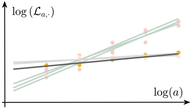


Finite differences $\mathbf{D}_1^{\rightarrow}\boldsymbol{x}$ (horizontal), $\mathbf{D}_1^{\uparrow}\boldsymbol{x}$ (vertical) at each pixel

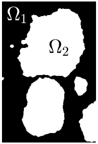
Functionals with either free or co-localized contours

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \quad \underbrace{\sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \quad + \quad \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model



→ favors piecewise constancy



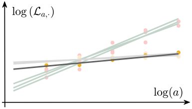
Finite differences $\mathbf{D}_1\boldsymbol{x} = \left[\mathbf{D}_1^{\rightarrow}\boldsymbol{x}, \mathbf{D}_1^{\uparrow}\boldsymbol{x}\right]$

Free: $\boldsymbol{h}$, $\boldsymbol{v}$ are **independently** piecewise constant

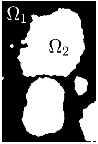
$$\mathcal{Q}_{\mathrm{F}}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha) = \alpha\|\mathbf{D}_1\boldsymbol{h}\|_{2,1} + \|\mathbf{D}_1\boldsymbol{v}\|_{2,1}$$

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \; \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \; + \; \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

→ fidelity to the log-linear model



→ favors piecewise constancy



Finite differences $\mathbf{D}_1\boldsymbol{x} = \left[\mathbf{D}_1^{\rightarrow}\boldsymbol{x}, \mathbf{D}_1^{\uparrow}\boldsymbol{x}\right]$

Free: $\boldsymbol{h}$, $\boldsymbol{v}$ are **independently** piecewise constant

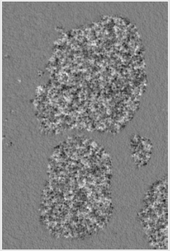
$$\mathcal{Q}_{\mathsf{F}}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha) = \alpha\|\mathbf{D}_1\boldsymbol{h}\|_{2,1} + \|\mathbf{D}_1\boldsymbol{v}\|_{2,1}$$

Co-localized: $\boldsymbol{h}$, $\boldsymbol{v}$ are **concomitantly** piecewise constant

$$\mathcal{Q}_{\mathsf{C}}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha) = \|[\alpha\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}]\|_{2,1}$$

$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \quad \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \quad + \quad \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image

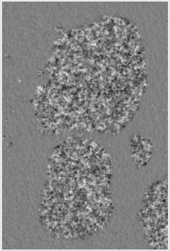


Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Segmentation *via* iterated thresholding

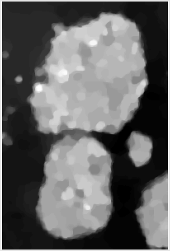
$$\underset{\boldsymbol{h},\boldsymbol{v}}{\text{minimize}} \ \sum_a \underbrace{\|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2}_{\textbf{Least-Squares}} \ + \ \lambda \underbrace{\mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)}_{\textbf{Total Variation}}$$

Textured image

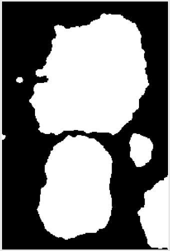


Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Co-localized contours $\widehat{\boldsymbol{h}}^{\mathrm{C}}$



Threshold estimate† $T\widehat{\boldsymbol{h}}^{\mathrm{C}}$



†(Cai et al., 2013, *J. Sci. Comput.*)

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

$(\lambda, \alpha) = (0, 0)$

Regularization parameters selection

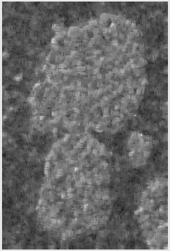
$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$

$(\lambda, \alpha) = (0.5, 0.5)$



too small

Regularization parameters selection

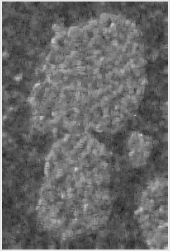
$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$

Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

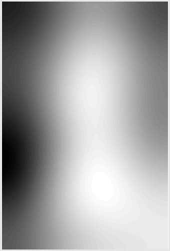
$(\lambda, \alpha) = (0, 0)$

$(\lambda, \alpha) = (0.5, 0.5)$



too small

$(\lambda, \alpha) = (500, 500)$



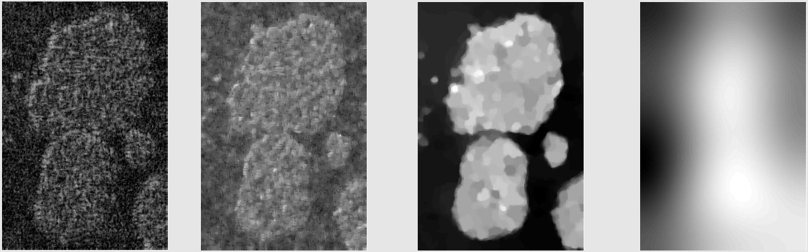
too large

Regularization parameters selection

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)$$

Lin. reg. $\widehat{\boldsymbol{h}}^{\mathrm{LR}}$ — Co-localized contours estimate $\widehat{\boldsymbol{h}}^{\mathrm{C}}$

$(\lambda, \alpha) = (0, 0)$ | $(\lambda, \alpha) = (0.5, 0.5)$ | $(\lambda^\dagger, \alpha^\dagger) = (11.5, 0.8)$ | $(\lambda, \alpha) = (500, 500)$



too small | optimal | too large

What *optimal* means? How to determine $\lambda^\dagger$ and $\alpha^\dagger$?

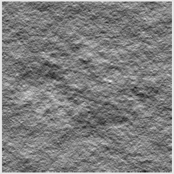
Parameter tuning (Grid search)

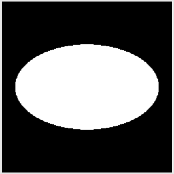
$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\mathcal{L}; \lambda, \alpha) = \operatorname*{argmin}_{\boldsymbol{h}, \boldsymbol{v}} \sum_{a} \|\log \mathcal{L}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \| \log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v} \|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Example





Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$

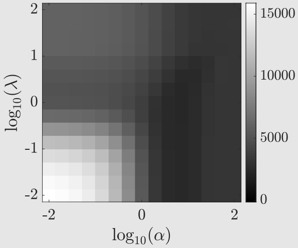
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$



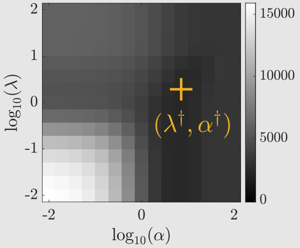
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\mathcal{L}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \mathcal{L}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\mathcal{L}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$

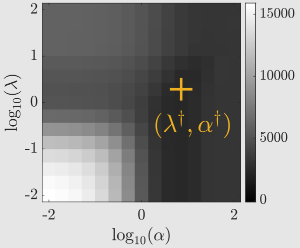


$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

?

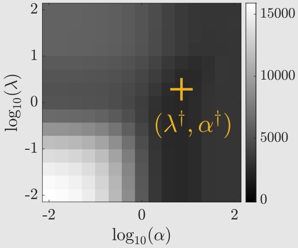
Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_a \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

$\boldsymbol{h}$: *discriminant*, $\boldsymbol{v}$: *auxiliary*

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

?

Stein Unbiased Risk Estimate
(SURE)

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\zeta \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

Ex. $$\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \underset{\boldsymbol{x}}{\operatorname{argmin}} \, \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

$$\textbf{Ex.} \quad \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \operatorname*{argmin}_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Quadratic error $R(\lambda) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \bar{\boldsymbol{x}}\|^2 \overset{?}{=} \mathbb{E}_{\boldsymbol{\zeta}} \widehat{R}(\boldsymbol{y}; \lambda)$ $\bar{\boldsymbol{x}}$ unknown

Observations $\boldsymbol{y} = \bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}}$: truth and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$

Parametric estimator $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$

Ex. $$\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \begin{cases} \left(\mathbf{I} + \lambda \mathbf{D}^\top \mathbf{D}\right)^{-1} \boldsymbol{y} & \text{(linear)} \\ \underset{\boldsymbol{x}}{\operatorname{argmin}} \|\boldsymbol{y} - \boldsymbol{x}\|^2 + \lambda \mathcal{Q}(\mathbf{D}\boldsymbol{x}) & \text{(nonlinear)} \end{cases}$$

Quadratic error $R(\lambda) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \bar{\boldsymbol{x}}\|^2 \overset{?}{=} \mathbb{E}_{\boldsymbol{\zeta}} \widehat{R}(\boldsymbol{y}; \lambda)$ $\bar{\boldsymbol{x}}$ unknown

Theorem (Stein, 1981,*Ann. Stat.*)

Let $(\boldsymbol{y}; \lambda) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)$ an estimator of $\bar{\boldsymbol{x}}$

- weakly differentiable w.r.t. $\boldsymbol{y}$,
- such that $\boldsymbol{\zeta} \mapsto \langle \widehat{\boldsymbol{x}}(\bar{\boldsymbol{x}} + \boldsymbol{\zeta}; \lambda), \boldsymbol{\zeta} \rangle$ is integrable w.r.t. $\mathcal{N}(\mathbf{0}, \rho^2 \mathbf{I})$.

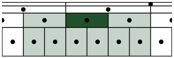
$$\widehat{R}(\boldsymbol{y}; \lambda) \triangleq \|\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) - \boldsymbol{y}\|^2 + 2\rho^2 \operatorname{tr}\left(\partial_{\boldsymbol{y}} \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda)\right) - \rho^2 P$$
$$\Longrightarrow R(\lambda) = \mathbb{E}_{\boldsymbol{\zeta}}[\widehat{R}(\boldsymbol{y}; \lambda)].$$

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\quad$ $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \mathcal{S})$ $\quad$ $\mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

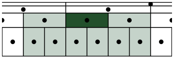
Projected estimation error $\quad R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Observations $\boldsymbol{y} = \boldsymbol{\Phi}\bar{\boldsymbol{x}} + \boldsymbol{\zeta} \in \mathbb{R}^P$, $\bar{\boldsymbol{x}} \in \mathbb{R}^N$, $\boldsymbol{\Phi} : \mathbb{R}^{P \times N}$ and $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$

E.g. the estimators $\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha)$ with free or co-localized contours

$\log \boldsymbol{\mathcal{L}} = \boldsymbol{\Phi}(\bar{\boldsymbol{h}}, \bar{\boldsymbol{v}}) + \boldsymbol{\zeta}$ $\qquad$ $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$ $\qquad$ $\mathcal{R} = \|\widehat{\boldsymbol{h}} - \bar{\boldsymbol{h}}\|^2$

$\boldsymbol{\Phi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto \{\log(a)\boldsymbol{h} + \boldsymbol{v}\}_a$



$\boldsymbol{\Pi} : (\boldsymbol{h}, \boldsymbol{v}) \mapsto (\boldsymbol{h}, \mathbf{0})$

Projected estimation error $\quad R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) \triangleq \mathbb{E}_{\boldsymbol{\zeta}} \|\boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{\Pi}\bar{\boldsymbol{x}}\|^2$

Theorem (Pascal et al., 2021, *J. Math. Imaging Vis.*)

Let $(\boldsymbol{y}; \boldsymbol{\Lambda}) \mapsto \widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})$ be an estimator of $\bar{\boldsymbol{x}}$

- weakly differentiable w.r.t. $\boldsymbol{y}$,
- such that $\boldsymbol{\zeta} \mapsto \langle \boldsymbol{\Pi}\widehat{\boldsymbol{x}}(\bar{\boldsymbol{x}} + \boldsymbol{\zeta}; \lambda), \mathbf{A}\boldsymbol{\zeta}\rangle$ is integrable w.r.t. $\mathcal{N}(\mathbf{0}, \boldsymbol{\mathcal{S}})$.

$$\widehat{R}(\boldsymbol{\Lambda}) \triangleq \|\mathbf{A}(\boldsymbol{\Phi}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda}) - \boldsymbol{y})\|^2 + 2\mathrm{tr}\left(\boldsymbol{\mathcal{S}}\mathbf{A}^\top \boldsymbol{\Pi}\partial_{\boldsymbol{y}}\widehat{\boldsymbol{x}}(\boldsymbol{y}; \boldsymbol{\Lambda})\right) - \mathrm{tr}\left(\mathbf{A}\boldsymbol{\mathcal{S}}\mathbf{A}^\top\right)$$

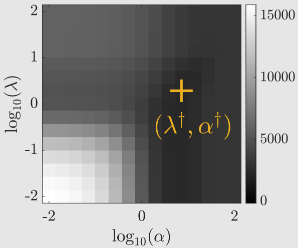
$$\Longrightarrow R_{\boldsymbol{\Pi}}(\boldsymbol{\Lambda}) = \mathbb{E}_{\boldsymbol{\zeta}}[\widehat{R}(\boldsymbol{\Lambda})].$$

Parameter tuning (Grid search)

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)$$

$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



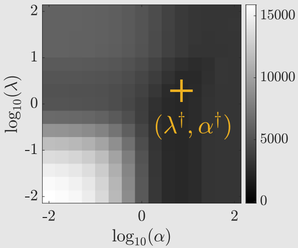
$\bar{\boldsymbol{h}}$: unknown!

$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \left\| \log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v} \right\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

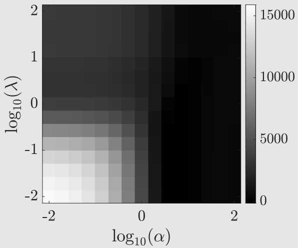
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

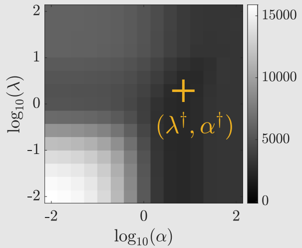
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$



$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1\boldsymbol{h}, \mathbf{D}_1\boldsymbol{v}; \alpha)$$

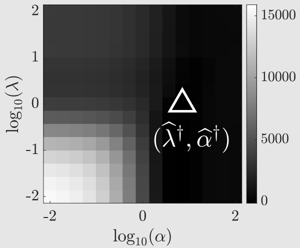
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\|\widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}}\right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

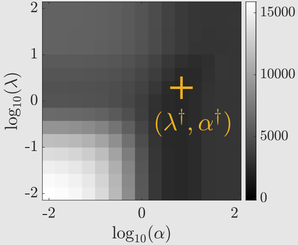
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$



$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h},\boldsymbol{v}}{\operatorname{argmin}} \sum_a \left\| \log \boldsymbol{\mathcal{L}}_{a,.} - \log(a)\boldsymbol{h} - \boldsymbol{v} \right\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

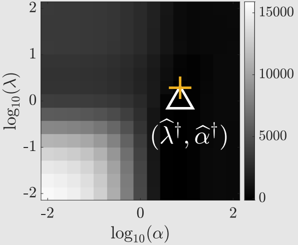
$\bar{\boldsymbol{h}}$: *true* regularity

$$\mathcal{R}(\lambda, \alpha) = \left\| \widehat{\boldsymbol{h}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha) - \bar{\boldsymbol{h}} \right\|^2$$



$\bar{\boldsymbol{h}}$: unknown!

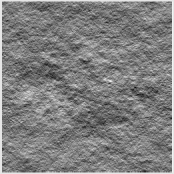
$$\widehat{R}_{\nu,\boldsymbol{\varepsilon}}(\boldsymbol{\mathcal{L}}; \lambda, \alpha | \boldsymbol{\mathcal{S}})$$

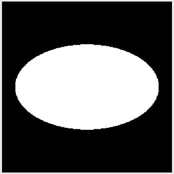


Parameter tuning (Grid search)

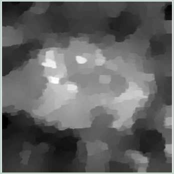
$$\left(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{v}}\right)(\boldsymbol{\mathcal{L}}; \lambda, \alpha) = \underset{\boldsymbol{h}, \boldsymbol{v}}{\operatorname{argmin}} \sum_{a} \|\log \boldsymbol{\mathcal{L}}_{a,\cdot} - \log(a)\boldsymbol{h} - \boldsymbol{v}\|^2 + \lambda \mathcal{Q}(\mathbf{D}_1 \boldsymbol{h}, \mathbf{D}_1 \boldsymbol{v}; \alpha)$$

Example

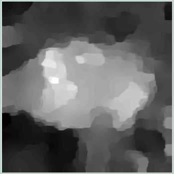




$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \lambda^{\dagger}, \alpha^{\dagger})$
(grid)

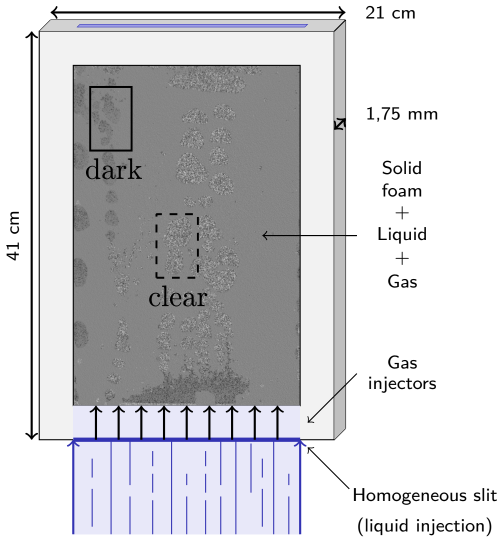


$\widehat{\boldsymbol{h}}^{\mathrm{F}}(\boldsymbol{\mathcal{L}}; \widehat{\lambda}^{\dagger}, \widehat{\alpha}^{\dagger})$
(grid)



Multiphase flow through porous media

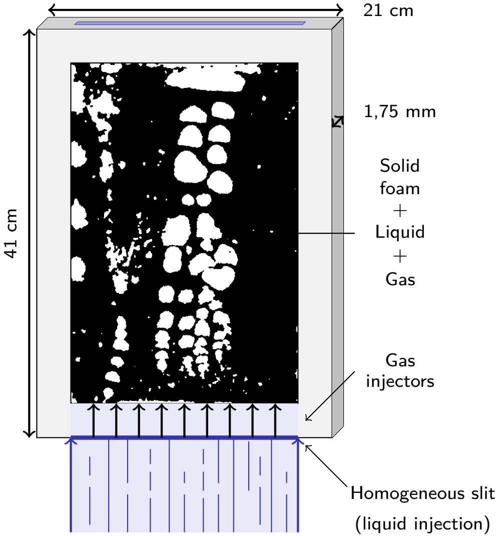
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



- → 1600 × 1100 pixels
- → video: ∼ 1000 images
- → phase diagram: ∼ 10 flow rates

Multiphase flow through porous media

Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)

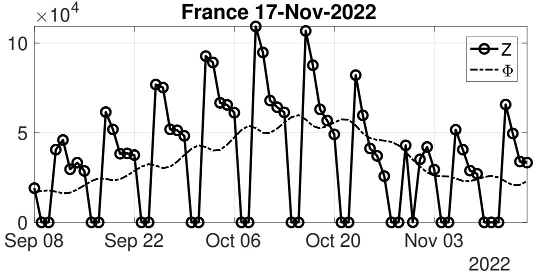


→ 1600 × 1100 pixels
→ video: ∼ 1000 images
→ phase diagram: ∼ 10 flow rates

Time series analysis:

Epidemiological indicator estimation

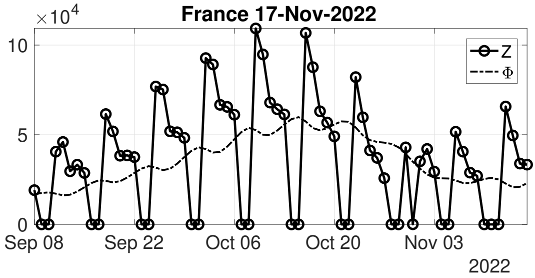
Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

$\Longrightarrow$ number of cases not informative enough: need to capture the **dynamics**

Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

$\Longrightarrow$ number of cases not informative enough: need to capture the **dynamics**

Design adapted counter measures and evaluate their effectiveness

$\longrightarrow$ efficient monitoring tools *epidemiological model,*

$\longrightarrow$ robust to low quality of the data *managing erroneous counts.*

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

Interpretation: at day t

$R_t > 1$ the virus propagates at exponential speed,

$R_t < 1$ the epidemic shrinks with an exponential decay,

$R_t = 1$ the epidemic is stable.

$\Longrightarrow$ one single indicator accounting for the overall pandemic mechanism

Reproduction number in Cori model

"averaged number of secondary cases generated by a typical infectious individual"

(Cori et al., 2013, *Am. Journal of Epidemiology*; Liu et al., 2018, *PNAS*)

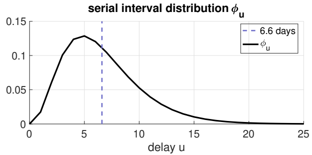
Interpretation: at day t

$\mathrm{R}_t > 1$ the virus propagates at exponential speed,
$\mathrm{R}_t < 1$ the epidemic shrinks with an exponential decay,
$\mathrm{R}_t = 1$ the epidemic is stable.

$\Longrightarrow$ one single indicator accounting for the overall pandemic mechanism

Principle: Z_t new infections at day t

$$\mathbb{E}\left[\mathrm{Z}_t\right] = \mathrm{R}_t \Phi_t, \quad \Phi_t = \sum_{u=1}^{\tau_\Phi} \phi_u \mathrm{Z}_{t-u}$$



with Φ_t global "infectiousness" in the population

$\{\phi_u\}_{u=1}^{\tau_\Phi}$ distribution of delay between onset of symptoms in primary and secondary cases

Gamma distribution truncated at 25 days, of mean 6.6 days and standard deviation 3.5 days

Pandemic study: modeling at the service of monitoring

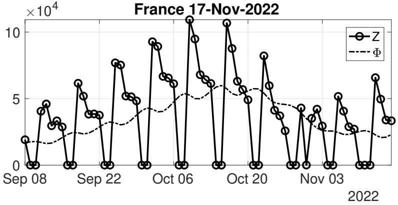
Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: **scaled** Poisson distribution

$$\frac{Z_t}{\gamma} \Big| Z_{t-\tau_\Phi}, \ldots, Z_{t-1}; R_t \sim \mathcal{P}\left(\frac{R_t \Phi_t}{\gamma}\right)$$

$\gamma > 0$ scaling parameter: controls **variance**

(Pascal & Vaiter, 2025, *Signal Process.*)



Pandemic study: modeling at the service of monitoring

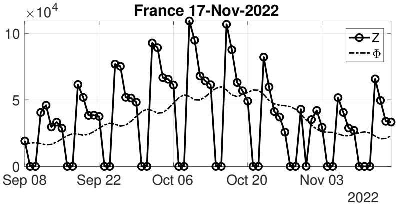
Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: **scaled** Poisson distribution

$$\frac{Z_t}{\gamma}\Big|Z_{t-\tau_\Phi}, \ldots, Z_{t-1}; R_t \sim \mathcal{P}\left(\frac{R_t\Phi_t}{\gamma}\right)$$

$\gamma > 0$ scaling parameter: controls **variance**

(Pascal & Vaiter, 2025, *Signal Process.*)



Inverse problem formalism: $\quad \frac{\mathbf{Z}}{\gamma} \sim \mathcal{P}\left(\frac{\mathbf{\Phi R}}{\gamma}\right)$

- $\mathbf{Z} \in \mathbb{N}^T$: reported infection counts,
- $\mathbf{R} = (R_1, \ldots, R_T) \in \mathbb{R}_+^T$: daily unknown reproduction number,
- $\mathbf{\Phi} = \mathrm{diag}\,(\Phi_1, \ldots, \Phi_T)$: linear operator,
- $\mathcal{P}$: data-dependent Poisson noise, with scale parameter γ

$$\Longrightarrow \mathcal{D}(\mathbf{Z}, \mathbf{\Phi R}) = -\log \mathbb{P}(\mathbf{Z}|\mathbf{R})$$

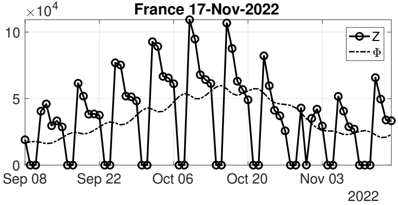
Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: **scaled** Poisson distribution

$$\frac{Z_t}{\gamma} \Big| Z_{t-\tau_\Phi}, \ldots, Z_{t-1}; R_t \sim \mathcal{P}\left(\frac{R_t \Phi_t}{\gamma}\right)$$

$\gamma > 0$ scaling parameter: controls **variance**

(Pascal & Vaiter, 2025, *Signal Process.*)



Maximum Likelihood Estimate (MLE)

$$\ln\left(\mathbb{P}(Z_t|\mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right)$$

$$\underset{\text{(def.)}}{=} -\frac{1}{\gamma} d_{\mathrm{KL}}(Z_t|R_t\Phi_t) \quad \text{(Kullback-Leibler)}$$

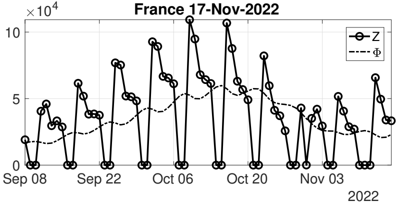
Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: **scaled** Poisson distribution

$$\frac{Z_t}{\gamma}\Big| Z_{t-\tau_\Phi}, \ldots, Z_{t-1}; R_t \sim \mathcal{P}\left(\frac{R_t \Phi_t}{\gamma}\right)$$

$\gamma > 0$ scaling parameter: controls **variance**

(Pascal & Vaiter, 2025, *Signal Process.*)



Maximum Likelihood Estimate (MLE)

$$\ln\left(\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right)$$

$$\underset{\text{(def.)}}{=} -\frac{1}{\gamma} d_{KL}(Z_t | R_t \Phi_t) \quad \text{(Kullback-Leibler)}$$

$$\Longrightarrow \widehat{R}_t^{\text{MLE}} = Z_t / \Phi_t = Z_t / \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$

ratio of moving averages

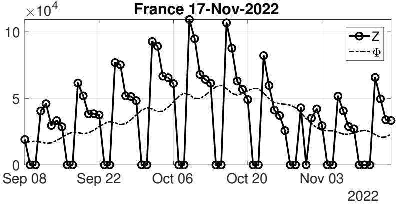
Data: daily counts $\mathbf{Z} = (Z_1, \ldots, Z_T)$

Model: **scaled** Poisson distribution

$$\frac{Z_t}{\gamma} \Big| Z_{t-\tau_\Phi}, \ldots, Z_{t-1}; R_t \sim \mathcal{P}\left(\frac{R_t \Phi_t}{\gamma}\right)$$

$\gamma > 0$ scaling parameter: controls **variance**

(Pascal & Vaiter, 2025, *Signal Process.*)



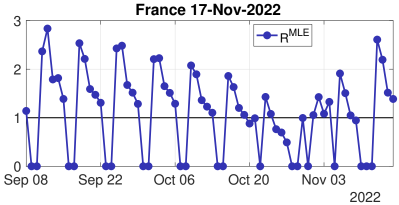
Maximum Likelihood Estimate (MLE)

$$\ln\left(\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)\right)$$

$$\underset{\text{(def.)}}{=} -\frac{1}{\gamma} d_{\mathrm{KL}}(Z_t | R_t \Phi_t) \quad \text{(Kullback-Leibler)}$$

$$\Longrightarrow \widehat{R}_t^{\mathrm{MLE}} = Z_t / \Phi_t = Z_t / \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$

ratio of moving averages



– huge variability along time/ no local trend

– not robust to pseudo-periodicity/ misreported counts

Penalized likelihood: regularization through nonlinear filtering

$$\widehat{\mathbf{R}}^{\mathrm{PKL}} = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{\gamma} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{R}_t \Phi_t\right) + \lambda \mathcal{R}(\mathbf{R}) \quad \text{(penalized Kullback-Leibler)}$$

with $\mathcal{R}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PLOSOne*)

Penalized likelihood: regularization through nonlinear filtering

$$\widehat{\mathbf{R}}^{\mathsf{PKL}} = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{\gamma} \mathsf{d}_{\mathsf{KL}}\left(\mathsf{Z}_t \,|\, \mathsf{R}_t \Phi_t\right) + \lambda \mathcal{R}(\mathbf{R}) \quad \text{(penalized Kullback-Leibler)}$$

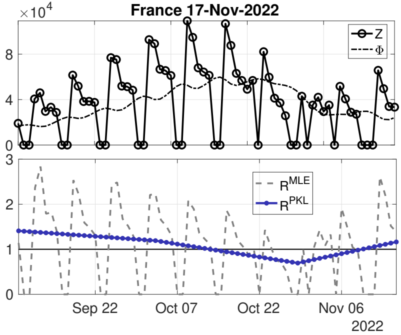
with $\mathcal{R}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PLOSOne*)

$$\mathcal{R}(\mathbf{R}) = \|\mathbf{D}_2\mathbf{R}\|_1$$

$$(\mathbf{D}_2\mathbf{R})_t = \mathsf{R}_{t+1} - 2\mathsf{R}_t + \mathsf{R}_{t-1}$$

2nd order derivative & ℓ_1-norm

$\Longrightarrow$ piecewise linearity



captures global **trend**, **smooth** temporal behavior, **no pseudo-oscillations**

Penalized Kullback-Leibler estimator:

$$\widehat{\mathbf{R}}(\mathbf{Z};\lambda) = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{\gamma} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \mid \mathrm{R}_t \Phi_t\right) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1$$

with $\mathcal{R}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PLOSOne*)

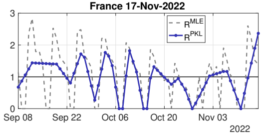
Penalized Kullback-Leibler estimator:

$$\widehat{\mathbf{R}}(\mathbf{Z};\lambda) = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{\gamma} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \mid \mathrm{R}_t \Phi_t\right) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1$$

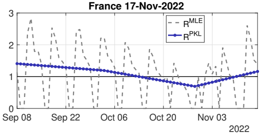
with $\mathcal{R}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PLOSOne*)

Fine tuning of the regularization parameter:

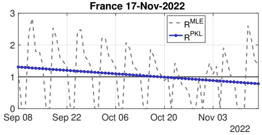
$\lambda = 3.5$



$\boldsymbol{\lambda^\dagger = 50}$



$\lambda = 250$



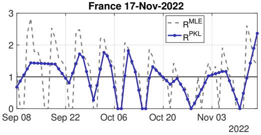
Penalized Kullback-Leibler estimator:

$$\widehat{\mathbf{R}}(\mathbf{Z};\lambda) = \underset{\mathbf{R}\in\mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^{T} \frac{1}{\gamma} \mathrm{d}_{\mathrm{KL}}\left(\mathrm{Z}_t \,|\, \mathrm{R}_t \Phi_t\right) + \lambda \|\mathbf{D}_2\mathbf{R}\|_1$$

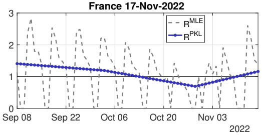
with $\mathcal{R}(\mathbf{R})$ favoring some temporal regularity (Abry et al., 2020, *PLOSOne*)

Fine tuning of the regularization parameter:

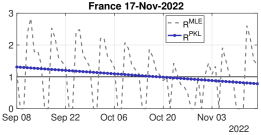
$\lambda = 3.5$



$\boldsymbol{\lambda}^\dagger = \mathbf{50}$



$\lambda = 250$



Data-driven oracle minimization

$$\boldsymbol{\lambda}^\dagger \in \underset{\lambda\in\Lambda}{\operatorname{Argmin}}\, \mathcal{O}(\mathbf{Z};\lambda)$$

$\Longrightarrow$ **Goal:** $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Goal: $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Strategy: Unbiased Risk Estimate $\mathbb{E}_{\mathbf{Z}}\left[\mathcal{O}(\mathbf{Z};\lambda)\right] = \mathbb{E}_{\mathbf{Z}}\left[\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2\right]$

Goal: $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Strategy: Unbiased Risk Estimate $\mathbb{E}_{\mathbf{Z}}\left[\mathcal{O}(\mathbf{Z};\lambda)\right] = \mathbb{E}_{\mathbf{Z}}\left[\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2\right]$

Reminder:

$$\frac{\mathbf{Z}}{\gamma} \sim \mathcal{P}\left(\frac{\mathbf{\Phi}\mathbf{R}}{\gamma}\right)$$

- $\mathbf{Z} \in \mathbb{N}^T$: reported infection counts,
- $\mathbf{R} = (\mathsf{R}_1, \dots, \mathsf{R}_T) \in \mathbb{R}_+^T$: daily unknown reproduction number,
- $\mathbf{\Phi} = \mathrm{diag}\,(\Phi_1, \dots, \Phi_T)$: linear operator,
- $\mathcal{P}$: data-dependent Poisson noise, with scale parameter γ.

Goal: $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Strategy: Unbiased Risk Estimate $\mathbb{E}_{\mathbf{Z}}\left[\mathcal{O}(\mathbf{Z};\lambda)\right] = \mathbb{E}_{\mathbf{Z}}\left[\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2\right]$

Reminder:
$$\frac{\mathbf{Z}}{\gamma} \sim \mathcal{P}\left(\frac{\mathbf{\Phi R}}{\gamma}\right)$$

- $\mathbf{Z} \in \mathbb{N}^T$: reported infection counts,
- $\mathbf{R} = (\mathrm{R}_1, \ldots, \mathrm{R}_T) \in \mathbb{R}_+^T$: daily unknown reproduction number,
- $\mathbf{\Phi} = \mathrm{diag}\,(\Phi_1, \ldots, \Phi_T)$: linear operator,
- $\mathcal{P}$: data-dependent Poisson noise, with scale parameter γ.

Challenges:

- Poisson model: Stein lemma does not apply (Eldar, 2008, *IEEE Trans. Signal Process.*; Luisier et al., 2010, *IEEE Trans. Image Process.*; Li et al., 2017, *IEEE Trans. Image Process.*)

Goal: $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Strategy: Unbiased Risk Estimate $\mathbb{E}_{\mathbf{Z}}\left[\mathcal{O}(\mathbf{Z};\lambda)\right] = \mathbb{E}_{\mathbf{Z}}\left[\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2\right]$

Reminder:

$$\frac{\mathbf{Z}}{\gamma} \sim \mathcal{P}\left(\frac{\boldsymbol{\Phi}(\mathbf{Z})\mathbf{R}}{\gamma}\right)$$

- $\mathbf{Z} \in \mathbb{N}^T$: reported infection counts,
- $\mathbf{R} = (\mathrm{R}_1, \ldots, \mathrm{R}_T) \in \mathbb{R}_+^T$: daily unknown reproduction number,
- $\boldsymbol{\Phi} = \mathrm{diag}\left(\Phi_1, \ldots, \Phi_T\right)$: linear operator,
- $\mathcal{P}$: data-dependent Poisson noise, with scale parameter γ.

Challenges:

- Poisson model: Stein lemma does not apply (Eldar, 2008, *IEEE Trans. Signal Process.*; Luisier et al., 2010, *IEEE Trans. Image Process.*; Li et al., 2017, *IEEE Trans. Image Process.*)
- Nonstationary driven autoregressive model: $(\boldsymbol{\Phi}\mathbf{R})_t = \mathrm{R}_t \sum_{s=1}^{\tau_\Phi} \phi_s \mathrm{Z}_{t-s}$

  $\Longrightarrow$ Novel counterpart of Stein lemma for driven autoregressive Poisson model

Goal: $\mathcal{O}$ data-driven proxy for $\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2$

Strategy: Unbiased Risk Estimate $\mathbb{E}_{\mathbf{Z}}\left[\mathcal{O}(\mathbf{Z};\lambda)\right] = \mathbb{E}_{\mathbf{Z}}\left[\|\widehat{\mathbf{R}}(\mathbf{Z};\lambda) - \overline{\mathbf{R}}\|_2^2\right]$

Reminder:

$$\frac{\mathbf{Z}}{\gamma} \sim \mathcal{P}\left(\frac{\boldsymbol{\Phi}(\mathbf{Z})\mathbf{R}}{\gamma}\right)$$

- $\mathbf{Z} \in \mathbb{N}^T$: reported infection counts,
- $\mathbf{R} = (\mathrm{R}_1, \ldots, \mathrm{R}_T) \in \mathbb{R}_+^T$: daily unknown reproduction number,
- $\boldsymbol{\Phi} = \mathrm{diag}\left(\Phi_1, \ldots, \Phi_T\right)$: linear operator,
- $\mathcal{P}$: data-dependent Poisson noise, with scale parameter γ.

Challenges:

- Poisson model: Stein lemma does not apply (Eldar, 2008, *IEEE Trans. Signal Process.*; Luisier et al., 2010, *IEEE Trans. Image Process.*; Li et al., 2017, *IEEE Trans. Image Process.*)
- Nonstationary driven autoregressive model: $(\boldsymbol{\Phi}\mathbf{R})_t = \mathrm{R}_t \sum_{s=1}^{\tau_\Phi} \phi_s \mathrm{Z}_{t-s}$

$\Longrightarrow$ Novel counterpart of Stein lemma for driven autoregressive Poisson model

Autoregressive **P**oisson **U**nbiased **R**isk **E**stimate (APURE)

Data-driven hyperparameter selection on synthetic Poisson data

Model

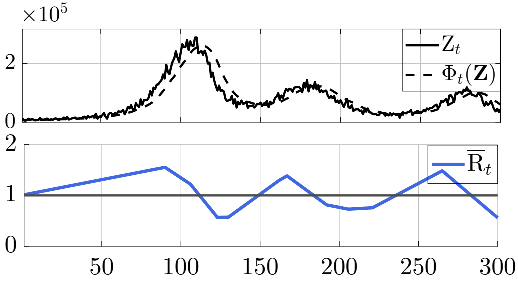
$$\frac{Z_t}{\gamma} \sim \mathcal{P}\left(\frac{\overline{R}_t \Phi_t(Z_1, \ldots, Z_{t-1})}{\gamma}\right)$$

with scale parameter $\log_{10} \gamma \equiv 3$

Model

$$\frac{Z_t}{\gamma} \sim \mathcal{P}\left(\frac{\overline{R}_t \Phi_t(Z_1, \ldots, Z_{t-1})}{\gamma}\right)$$

with scale parameter $\log_{10} \gamma \equiv 3$



Model

$$\frac{Z_t}{\gamma} \sim \mathcal{P}\left(\frac{\overline{R}_t \Phi_t(Z_1,\dots,Z_{t-1})}{\gamma}\right)$$

with scale parameter $\log_{10}\gamma \equiv 3$

True prediction $\mathcal{P}$ & estimation $\mathcal{E}$ errors

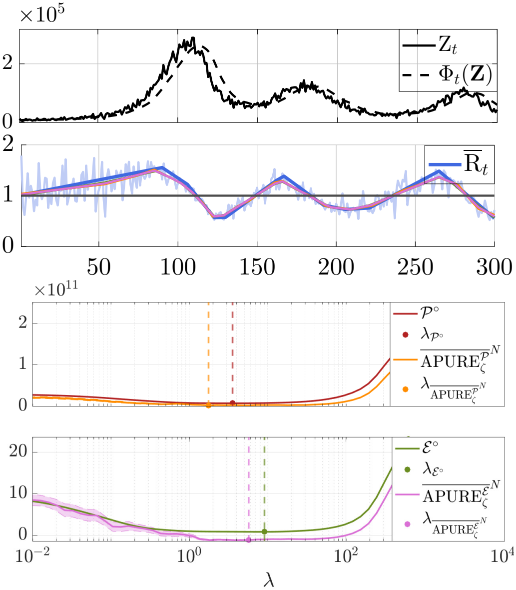
$$\mathcal{P}^\circ(\mathbf{Z};\boldsymbol{\theta}) = \|\widehat{\mathbf{R}}(\mathbf{Z};\boldsymbol{\theta}) \odot \boldsymbol{\Phi}(\mathbf{Z}) - \overline{\mathbf{R}} \odot \boldsymbol{\Phi}(\mathbf{Z})\|_2^2$$

$$\mathcal{E}^\circ(\mathbf{Z};\boldsymbol{\theta}) = \|\widehat{\mathbf{R}}(\mathbf{Z};\boldsymbol{\theta}) - \overline{\mathbf{R}}\|_2^2$$

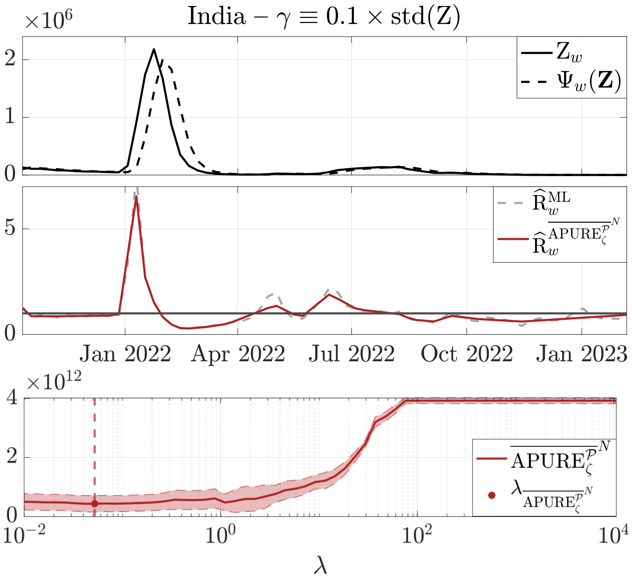
Unbiased risk estimators

$$\overline{\mathrm{APURE}_\zeta^{\mathcal{P}}}^N = \frac{1}{N}\sum_{n=1}^{N} \mathrm{APURE}_{\zeta^{(n)}}^{\mathcal{P}}$$

$$\overline{\mathrm{APURE}_\zeta^{\mathcal{E}}}^N = \frac{1}{N}\sum_{n=1}^{N} \mathrm{APURE}_{\zeta^{(n)}}^{\mathcal{E}}$$



Data-driven hyperparameter selection on weekly COVID-19 counts



Pascal & Vaiter, 2024, *Preprint arXiv:2409.14937*

Codes: github.com/bpascal-fr/APURE-Estim-Epi

Inverse problem $\boldsymbol{y} \sim \mathcal{B}\left(\boldsymbol{\Phi}\overline{\boldsymbol{x}}\right)$

$$\lambda^\dagger \in \underset{\lambda \in \Lambda}{\mathrm{Argmin}}\, \mathcal{O}(\boldsymbol{y}; \lambda), \quad \text{for} \quad \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) \in \underset{\boldsymbol{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\, \mathcal{D}(\boldsymbol{y}, \boldsymbol{\Phi}\boldsymbol{x}) + \lambda \mathcal{R}(\boldsymbol{x})$$

Inverse problem $\boldsymbol{y} \sim \mathcal{B}(\boldsymbol{\Phi}\overline{\boldsymbol{x}})$

$$\lambda^\dagger \in \underset{\lambda \in \Lambda}{\mathrm{Argmin}}\, \mathcal{O}(\boldsymbol{y};\lambda), \quad \text{for} \quad \widehat{\boldsymbol{x}}(\boldsymbol{y};\lambda) \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\mathrm{Argmin}}\, \mathcal{D}(\boldsymbol{y},\boldsymbol{\Phi}\boldsymbol{x}) + \lambda\mathcal{R}(\boldsymbol{x})$$

Data-driven parameter selection

$\Longrightarrow$ $\mathcal{O}$: Unbiased Risk Estimate (Stein, 1981, *Ann. Stat.*; Eldar, 2008, *IEEE Trans. Signal Process.*; Luisier et al., 2010, *IEEE Trans. Image Process.*; Deledalle et al., 2014, *SIAM J. Imaging Sci.*; Pascal et al., 2021, *J. Math. Imaging Vis.*; Lucas et al., 2023, *Signal, Image Video Process.*)

- Texture segmentation: additive correlated Gaussian noise;
- Epidemic monitoring: driven autoregressive data-dependent Poisson noise.

Inverse problem $\boldsymbol{y} \sim \mathcal{B}(\boldsymbol{\Phi}\overline{\boldsymbol{x}})$

$$\lambda^\dagger \in \underset{\lambda \in \Lambda}{\mathrm{Argmin}}\, \mathcal{O}(\boldsymbol{y}; \lambda), \quad \text{for} \quad \widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) \in \underset{\boldsymbol{x} \in \mathbb{R}^N}{\mathrm{Argmin}}\, \mathcal{D}(\boldsymbol{y}, \boldsymbol{\Phi}\boldsymbol{x}) + \lambda \mathcal{R}(\boldsymbol{x})$$

Data-driven parameter selection

$\Longrightarrow$ $\mathcal{O}$: Unbiased Risk Estimate (Stein, 1981, *Ann. Stat.*; Eldar, 2008, *IEEE Trans. Signal Process.*; Luisier et al., 2010, *IEEE Trans. Image Process.*; Deledalle et al., 2014, *SIAM J. Imaging Sci.*; Pascal et al., 2021, *J. Math. Imaging Vis.*; Lucas et al., 2023, *Signal, Image Video Process.*)

- Texture segmentation: additive correlated Gaussian noise;
- Epidemic monitoring: driven autoregressive data-dependent Poisson noise.

Extensions and perspectives

- Efficient and robust scheme for nonconvex $\mathcal{R}(\boldsymbol{x})$;
- Generalization to other noise models: speckle noise in medical imaging;
- Unsupervised learning for $\widehat{\boldsymbol{x}}(\boldsymbol{y}; \lambda) = \mathbf{NN}_{\boldsymbol{\theta}}(\boldsymbol{y})$ with loss $\mathcal{O}(\boldsymbol{y}; \boldsymbol{\theta})$.