



Laboratoire des Sciences  
du Numérique de Nantes



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

**Barbara Pascal**, [bpascal-fr.github.io](https://bpascal-fr.github.io)

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

**February 19, 2026**

**AI hackathon for women in the mathematical sciences**

ICMS, Edinburgh, UK

## Observation model

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

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

## Observation model

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

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

**Goal:** Estimate  $\bar{\mathbf{x}}$



## Observation model

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

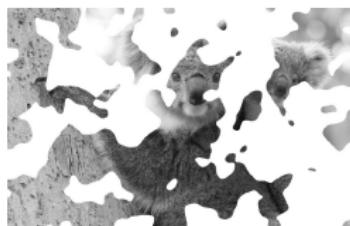
- $\mathbf{y} \in \mathbb{R}^P$ : degraded observations;
- $\bar{\mathbf{x}} \in \mathbb{R}^N$ : unknown quantity of interest;
- $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^P$ : known deformation;
- $\mathcal{B}$ : random measurement noise.

**Goal:** Estimate  $\bar{\mathbf{x}}$



► ill-conditioned, rank deficient  $\Phi$

Inpainting



Super-resolution



Deblurring



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

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

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

## Observation model

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

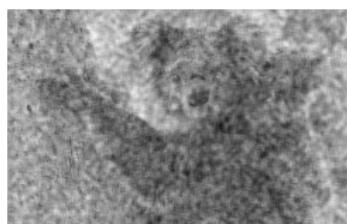
- $\mathbf{y} \in \mathbb{R}^P$ : degraded observations;
- $\bar{\mathbf{x}} \in \mathbb{R}^N$ : unknown quantity of interest;
- $\Phi : \mathbb{R}^N \rightarrow \mathbb{R}^P$ : known deformation;
- $\mathcal{B}$ : random measurement noise.

**Goal:** Estimate  $\bar{\mathbf{x}}$



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

Correlated



Data-dependent



Multiplicative



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

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

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

# The variational framework: penalized log-likelihood

## Variational estimator

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

- $\mathcal{D}(\mathbf{y}; \cdot) = -\log \mathbb{P}(\mathbf{y}|\cdot)$ : negative log-likelihood

Ex:  $\mathcal{D}(\mathbf{y}; \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2$

No regularization



$$\mathcal{R} = 0$$

# The variational framework: penalized log-likelihood

## Variational estimator

$$\hat{\mathbf{x}}(\mathbf{y}; \lambda) \in \operatorname{Argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{D}(\mathbf{y}, \Phi \mathbf{x}) + \lambda \mathcal{R}(\mathbf{x})$$

- $\mathcal{D}(\mathbf{y}; \cdot) = -\log \mathbb{P}(\mathbf{y}|\cdot)$ : negative log-likelihood
- $\mathcal{R}$ : regularization term encoding a priori knowledge

Ex:  $\mathcal{D}(\mathbf{y}; \mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2$

Ex:  $\mathcal{R}(\mathbf{x}) = \|\mathbf{D}_1 \mathbf{x}\|_q^q$

(Giovannelli & Idier, 2015, Wiley)

No regularization



$$\mathcal{R} = 0$$

Smooth



$$\mathcal{R}(\mathbf{x}) = \|\mathbf{D}_1 \mathbf{x}\|_2^2$$

(Tikhonov et al., 1977, Wiley)

Piecewise constant



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

(Rudin et al., 1992, Physica D)

## Hyperparameter selection: bilevel optimization

### Fine-tuning of the regularization parameter

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

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$$\lambda = 1$$



not enough

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$\lambda = 1$



not enough

$\lambda = 25$



too regularized

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$$\lambda = 1$$



not enough

$$\lambda^\dagger = 5$$



optimal

$$\lambda = 25$$



too regularized

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$$\lambda = 1$$



not enough

$$\lambda^\dagger = 5$$



optimal

$$\lambda = 25$$



too regularized

## Oracle-based hyperparameter selection

$$\lambda^\dagger \in \operatorname{Argmin}_{\lambda \in \Lambda} \mathcal{O}(\mathbf{y}; \lambda)$$

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$$\lambda = 1$$



not enough

$$\lambda^\dagger = 5$$



optimal

$$\lambda = 25$$



too regularized

## Oracle-based hyperparameter selection

$$\lambda^\dagger \in \operatorname{Argmin}_{\lambda \in \Lambda} \mathcal{O}(\mathbf{y}; \lambda)$$

- golden case:  $\mathcal{O}(\mathbf{y}; \lambda) = \|\hat{\mathbf{x}}(\mathbf{y}; \lambda) - \bar{\mathbf{x}}\|^2 \implies$  efficient bi-level  $(\mathbf{x}, \lambda)$  minimization

# Hyperparameter selection: bilevel optimization

## Fine-tuning of the regularization parameter

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

$$\lambda = 1$$



not enough

$$\lambda^\dagger = 5$$



optimal

$$\lambda = 25$$



too regularized

## Oracle-based hyperparameter selection

$$\lambda^\dagger \in \operatorname{Argmin}_{\lambda \in \Lambda} \mathcal{O}(\mathbf{y}; \lambda)$$

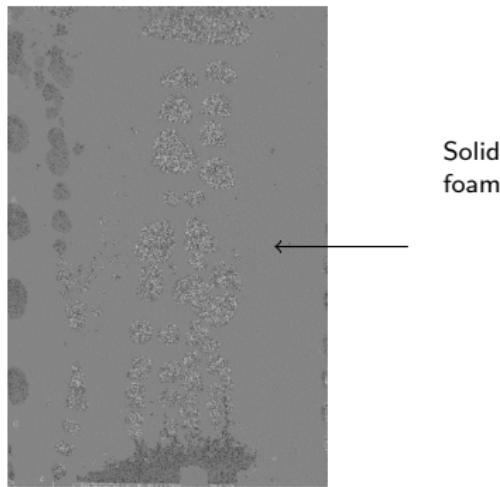
- golden case:  $\mathcal{O}(\mathbf{y}; \lambda) = \|\hat{\mathbf{x}}(\mathbf{y}; \lambda) - \bar{\mathbf{x}}\|^2 \implies$  efficient bi-level  $(\mathbf{x}, \lambda)$  minimization
- practical case: ground truth  $\bar{\mathbf{x}}$  **not available!**  $\implies$  data-driven  $\mathcal{O}(\mathbf{y};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

Image processing:

Texture segmentation

# Multiphase flow through porous media

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

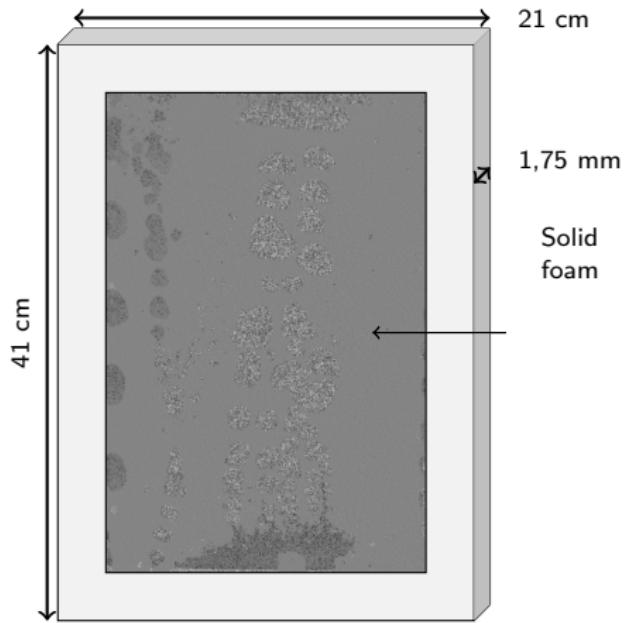


Solid  
foam



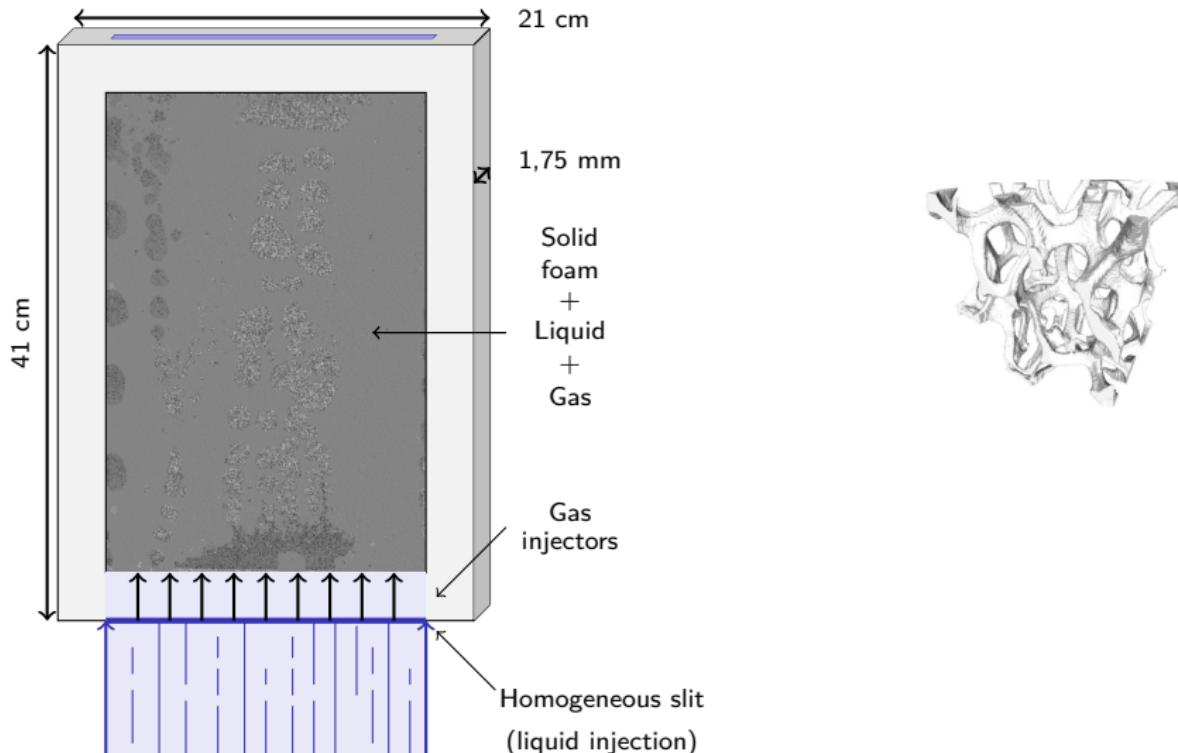
# Multiphase flow through porous media

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



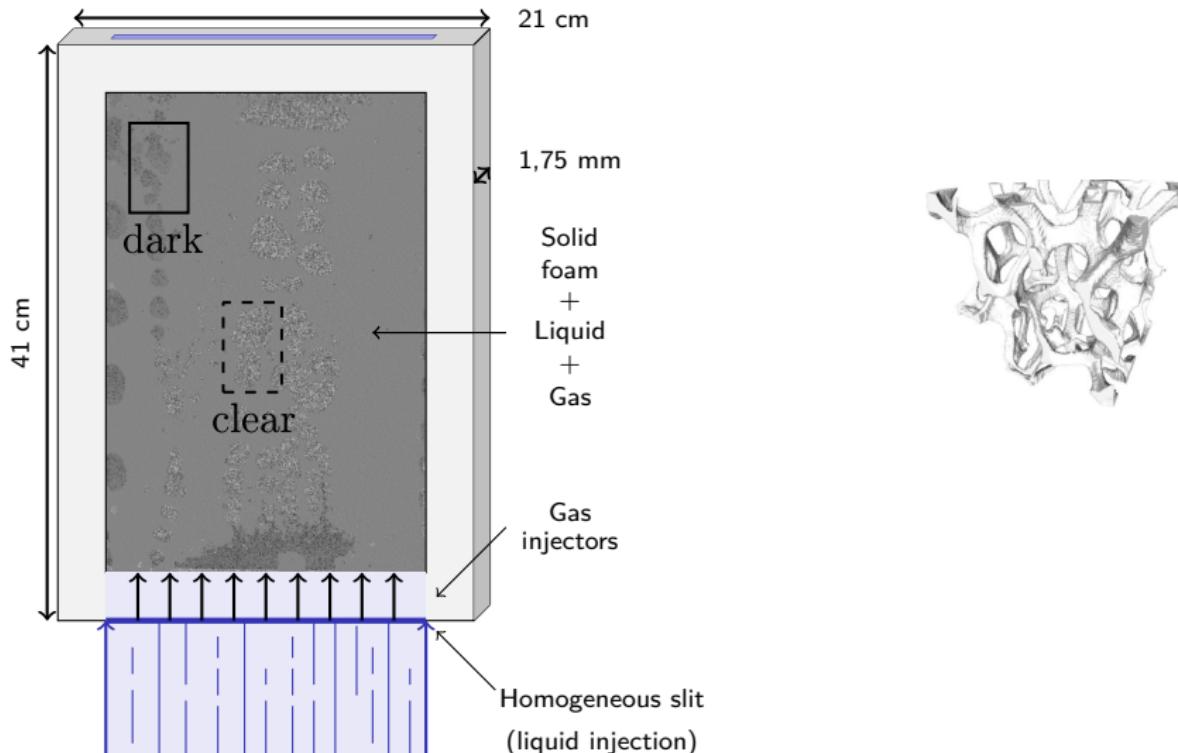
# Multiphase flow through porous media

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



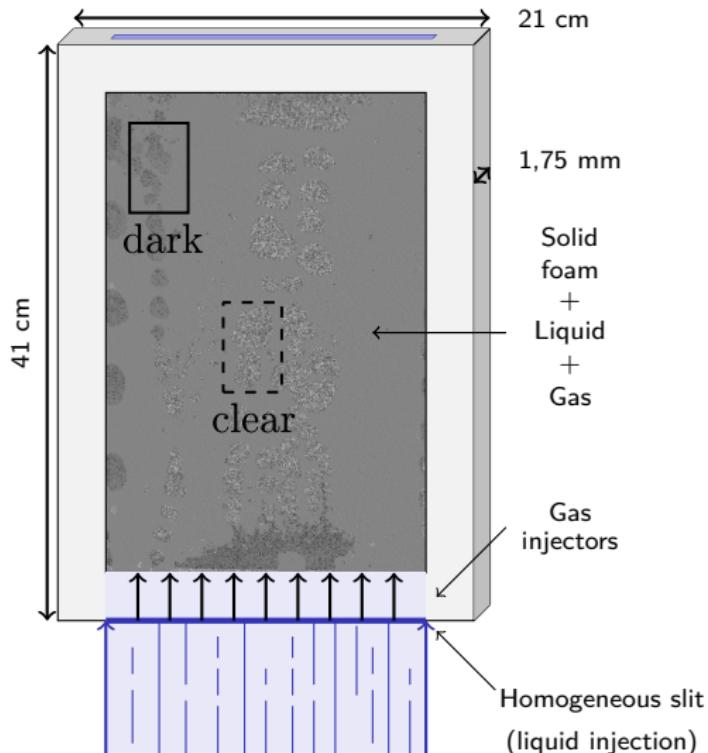
# Multiphase flow through porous media

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



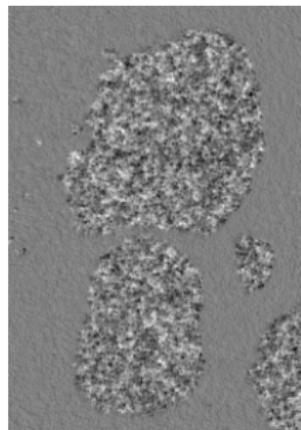
# Multiphase flow through porous media

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

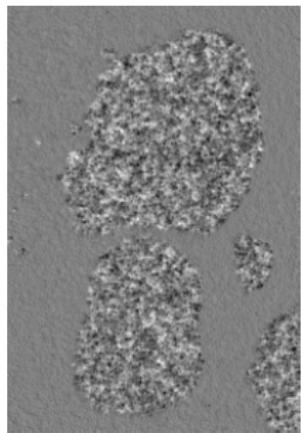


- 1600 × 1100 pixels
- video: ~ 1000 images
- phase diagram: ~ 10 flow rates

## Textured image segmentation



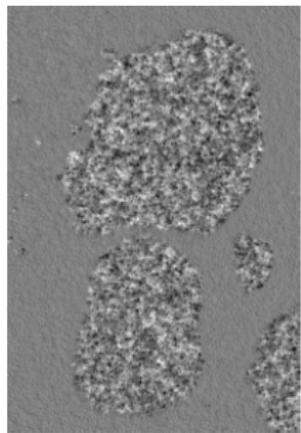
## Textured image segmentation



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

$$\Omega = \Omega_1 \sqcup \dots \sqcup \Omega_K$$

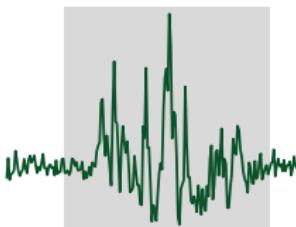
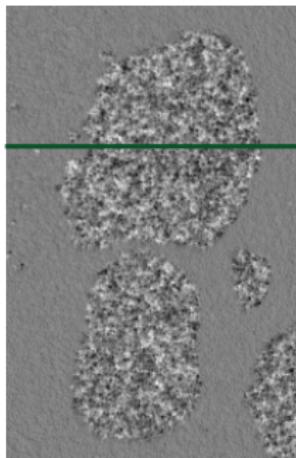
## Textured image segmentation



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

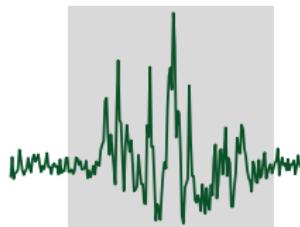
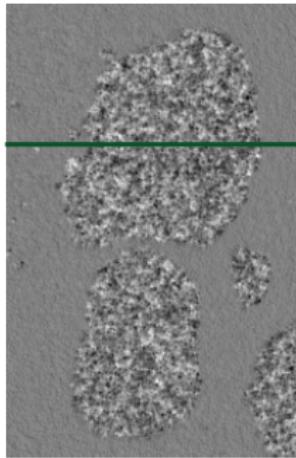
$$\Omega = \Omega_1 \sqcup \dots \sqcup \Omega_K$$

## Piecewise monofractal model



## Fractal attributes

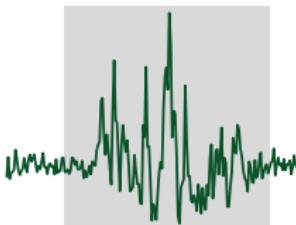
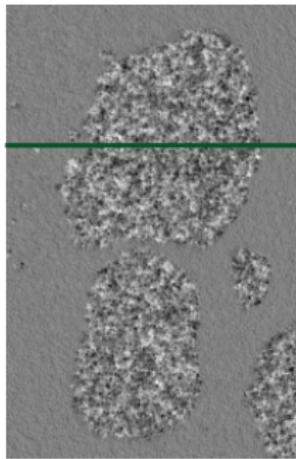
- variance  $\sigma^2$       *amplitude of variations*



# Piecewise monofractal model

## Fractal attributes

- variance  $\sigma^2$       *amplitude of variations*
- local regularity  $h$       *scale invariance*

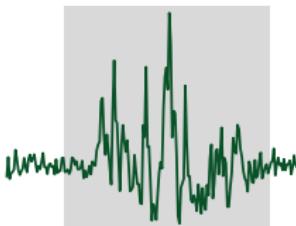
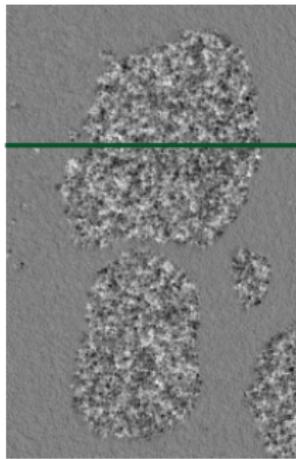


# Piecewise monofractal model

## Fractal attributes

- variance  $\sigma^2$       *amplitude of variations*
- local regularity  $h$       *scale invariance*

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



# Piecewise monofractal model

## Fractal attributes

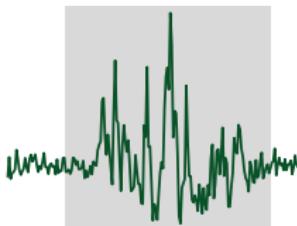
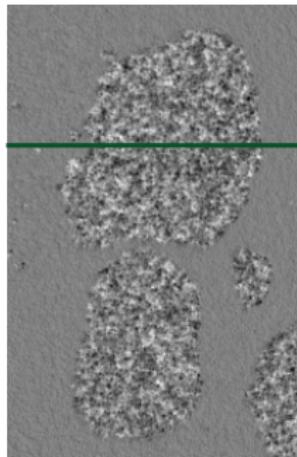
- variance  $\sigma^2$       *amplitude of variations*
- local regularity  $h$       *scale invariance*

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



$$h(x) \equiv h_1 = 0.9$$

$$h(x) \equiv h_2 = 0.3$$



# Piecewise monofractal model

## Fractal attributes

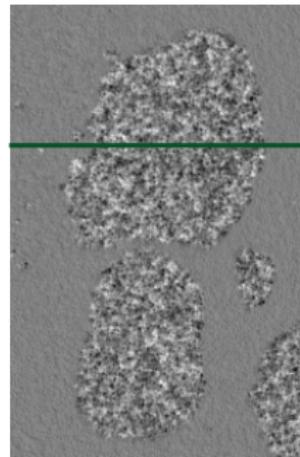
- variance  $\sigma^2$       *amplitude of variations*
- local regularity  $h$       *scale invariance*

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



$$h(x) \equiv h_1 = 0.9$$

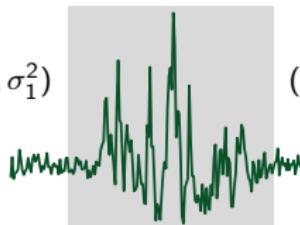
$$h(x) \equiv h_2 = 0.3$$



$$(h_2, \sigma_2^2)$$

$$(h_1, \sigma_1^2)$$

$$(h_1, \sigma_1^2)$$



## Segmentation

- $\sigma^2$  and  $h$  piecewise constant

# Piecewise monofractal model

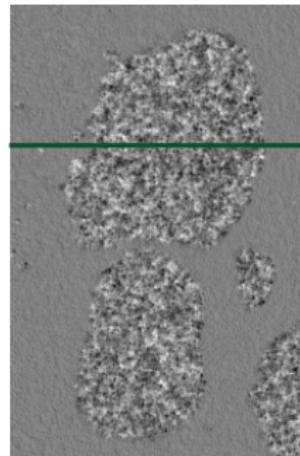
## Fractal attributes

- variance  $\sigma^2$       *amplitude of variations*
- local regularity  $h$       *scale invariance*

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

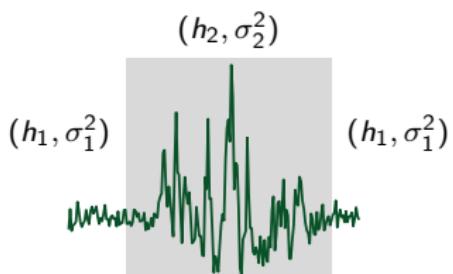


$$h(x) \equiv h_1 = 0.9 \quad h(x) \equiv h_2 = 0.3$$



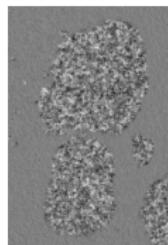
## Segmentation

- ▶  $\sigma^2$  and  $h$  piecewise constant
- ▶ region  $\Omega_k$  characterized by  $(\sigma_k^2, h_k)$



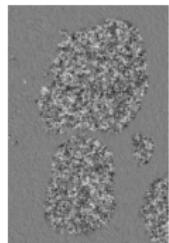
# Multiscale analysis

Textured image



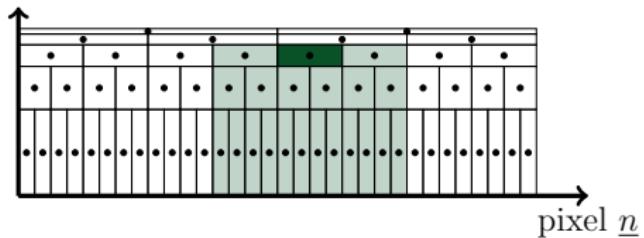
# Multiscale analysis

Textured image



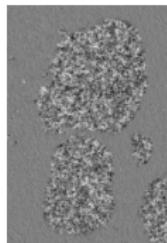
Local maximum of wavelet coefficients:  $\mathcal{L}_{a,\cdot}$

scale  $2^j$



# Multiscale analysis

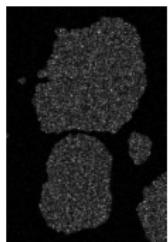
Textured image



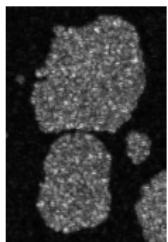
Local maximum of wavelet coefficients:  $\mathcal{L}_{a,\cdot}$

Scale

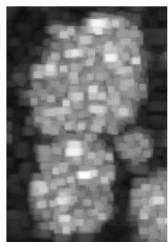
$a = 2^1$



$a = 2^2$

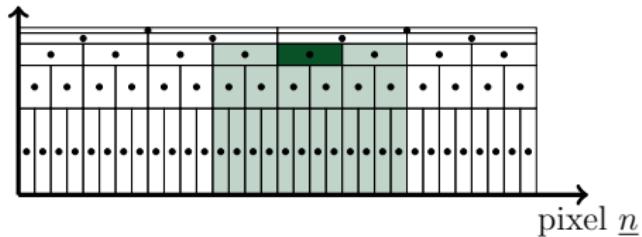


$a = 2^5$



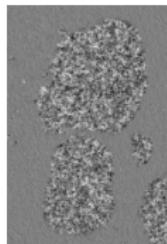
...

scale  $2^j$



# Multiscale analysis

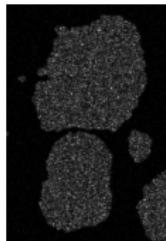
Textured image



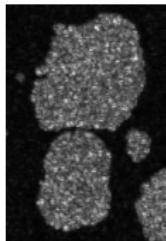
Local maximum of wavelet coefficients:  $\mathcal{L}_{a,\cdot}$

Scale

$a = 2^1$

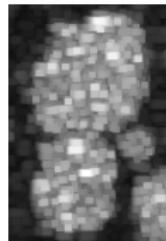


$a = 2^2$



...

$a = 2^5$



Proposition (Jaffard, 2004, *Proc. Symp. Pure*

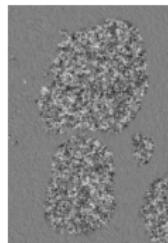
*Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log(\mathcal{L}_{a,\cdot}) \underset{a \rightarrow 0}{\simeq} \log(a) \frac{\mathbf{h}}{\text{regularity}} + \frac{\mathbf{v}}{\propto \log(\sigma^2)}$$

(variance)

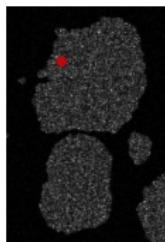
# Multiscale analysis

Textured image

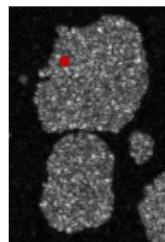


Scale

$a = 2^1$

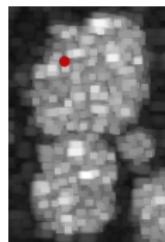


$a = 2^2$



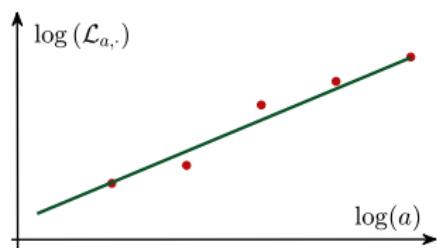
...

$a = 2^5$



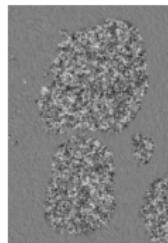
**Proposition** (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

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



# Multiscale analysis

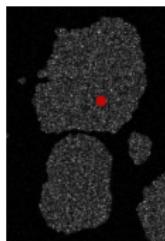
Textured image



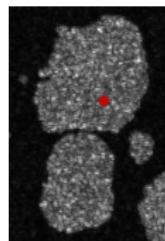
Local maximum of wavelet coefficients:  $\mathcal{L}_{a,\cdot}$

Scale

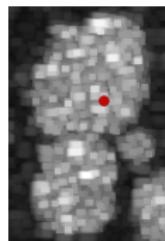
$a = 2^1$



$a = 2^2$



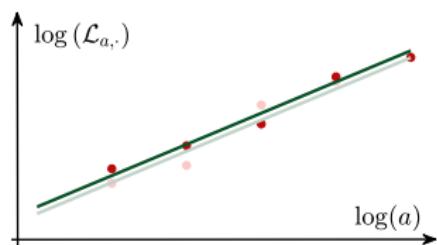
$a = 2^5$



...

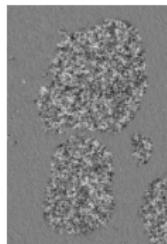
**Proposition** (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log(\mathcal{L}_{a,\cdot}) \underset{a \rightarrow 0}{\simeq} \log(a) h_{\text{regularity}} + v_{\propto \log(\sigma^2)} \text{ (variance)}$$



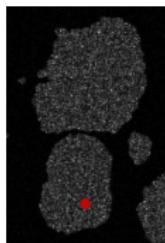
# Multiscale analysis

Textured image

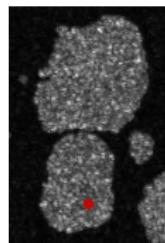


Scale

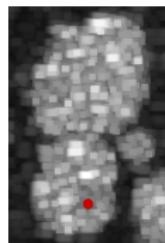
$a = 2^1$



$a = 2^2$



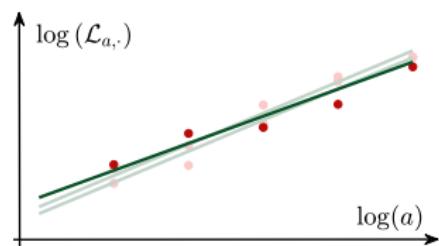
$a = 2^5$



...

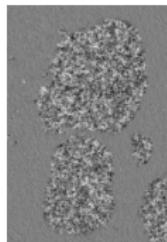
**Proposition** (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log(\mathcal{L}_{a,\cdot}) \underset{a \rightarrow 0}{\simeq} \log(a) h_{\text{regularity}} + v_{\propto \log(\sigma^2)} \text{ (variance)}$$



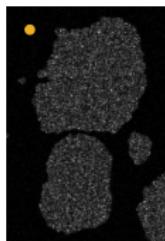
# Multiscale analysis

Textured image

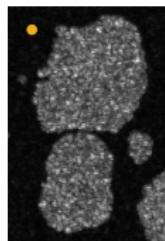


Scale

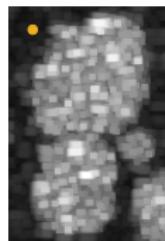
$a = 2^1$



$a = 2^2$



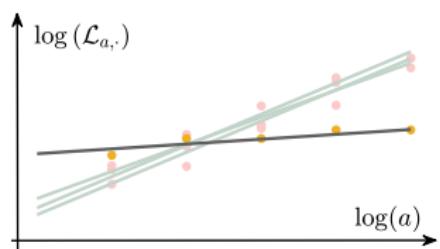
$a = 2^5$



...

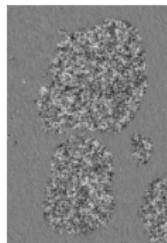
**Proposition** (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

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



# Multiscale analysis

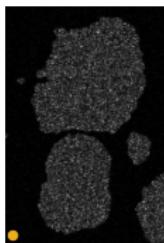
Textured image



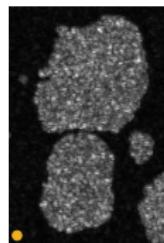
Local maximum of wavelet coefficients:  $\mathcal{L}_{a,\cdot}$

Scale

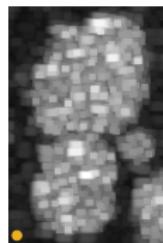
$a = 2^1$



$a = 2^2$



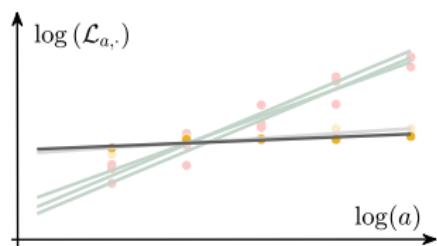
$a = 2^5$



...

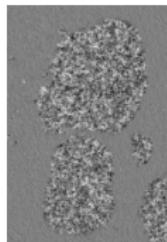
**Proposition** (Jaffard, 2004, *Proc. Symp. Pure Math.*; Wendt et al., 2009, *Signal Process.*)

$$\log(\mathcal{L}_{a,\cdot}) \underset{a \rightarrow 0}{\simeq} \log(a) h_{\text{regularity}} + v_{\propto \log(\sigma^2)} \text{ (variance)}$$



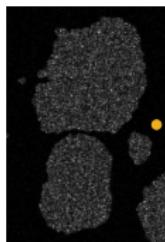
# Multiscale analysis

Textured image

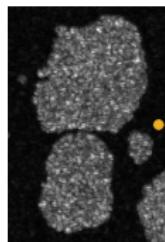


Scale

$a = 2^1$

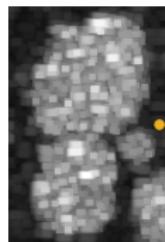


$a = 2^2$



...

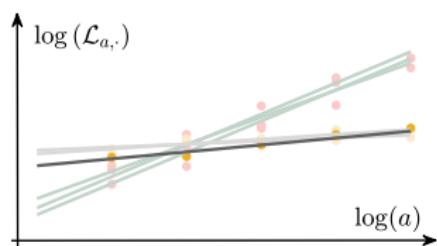
$a = 2^5$



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

$$\log(\mathcal{L}_{a,\cdot}) \underset{a \rightarrow 0}{\simeq} \log(a) h_{\text{regularity}} + v_{\propto \log(\sigma^2)}$$

(variance)

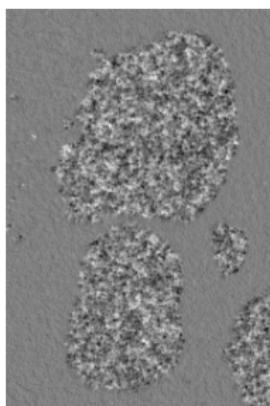


## Direct punctual estimation

**Linear regression**

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

Textured image

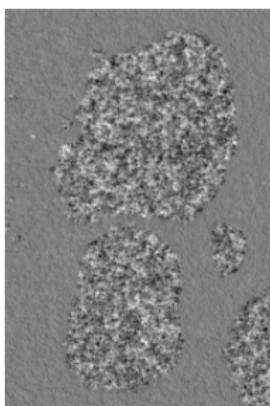


## Direct punctual estimation

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

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

Textured image

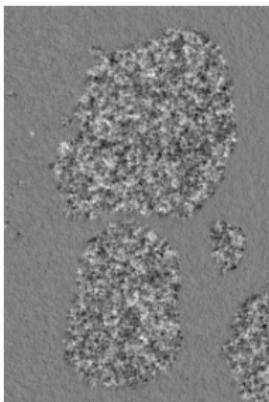


## Direct punctual estimation

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

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

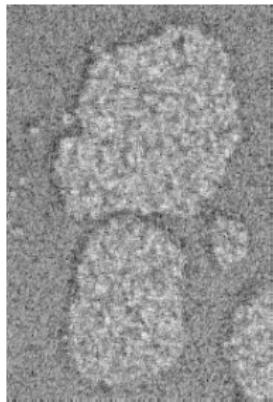
Textured image



Local regularity  $\hat{\boldsymbol{h}}^{\text{LR}}$



Local power  $\hat{\boldsymbol{v}}^{\text{LR}}$



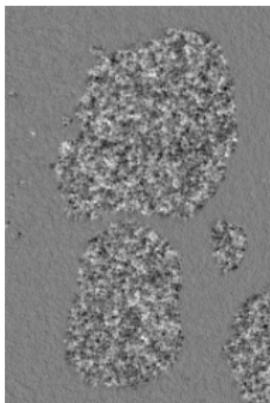
## Direct punctual estimation

### Linear regression

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

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

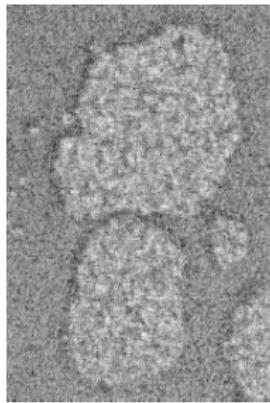
Textured image



Local regularity  $\hat{\boldsymbol{h}}^{\text{LR}}$



Local power  $\hat{\boldsymbol{v}}^{\text{LR}}$

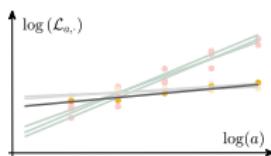


→ large estimation variance

## Functionals with either free or co-localized contours

$$\sum_a \frac{\|\log \mathcal{L}_{a,.} - \log(a) \mathbf{h} - \mathbf{v}\|^2}{\text{Least-Squares}}$$

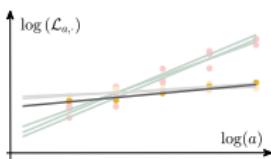
→ fidelity to the log-linear model



## Functionals with either free or co-localized contours

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

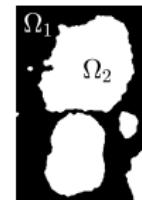
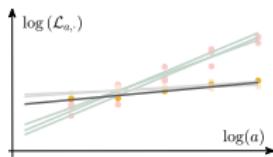
$\rightarrow$  fidelity to the log-linear model  
 $\rightarrow$  favors piecewise constancy



## Functionals with either free or co-localized contours

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

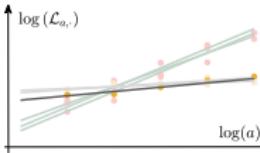
$\rightarrow$  fidelity to the log-linear model  
 $\rightarrow$  favors piecewise constancy



## Functionals with either free or co-localized contours

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

$\rightarrow$  fidelity to the log-linear model  
 $\rightarrow$  favors piecewise constancy

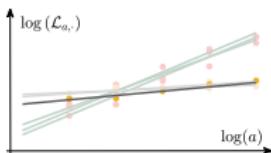


**Finite differences**  $\mathbf{D}_1^\rightarrow \mathbf{x}$  (horizontal),  $\mathbf{D}_1^\uparrow \mathbf{x}$  (vertical) at each pixel

## Functionals with either free or co-localized contours

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

$\rightarrow$  fidelity to the log-linear model  
 $\rightarrow$  favors piecewise constancy



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

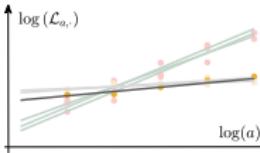
Free:  $\mathbf{h}$ ,  $\mathbf{v}$  are **independently** piecewise constant

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

## Functionals with either free or co-localized contours

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

$\rightarrow$  fidelity to the log-linear model  
 $\rightarrow$  favors piecewise constancy



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

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

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

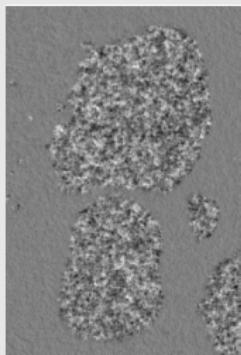
Co-localized:  $\mathbf{h}, \mathbf{v}$  are **concomitantly** piecewise constant

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

## Segmentation via iterated thresholding

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

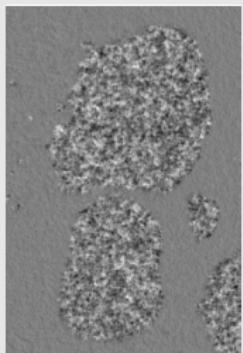
Textured image    Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$



## Segmentation via iterated thresholding

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

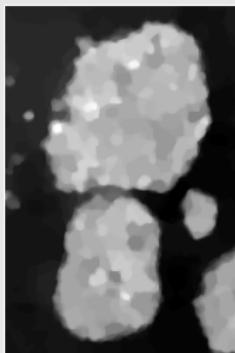
Textured image



Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$



Co-localized  
contours  $\hat{\mathbf{h}}^{\text{C}}$



Threshold  
estimate<sup>†</sup>  $T\hat{\mathbf{h}}^{\text{C}}$



<sup>†</sup>(Cai et al., 2013, *J. Sci. Comput.*)

## Regularization parameters selection

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

## Regularization parameters selection

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

Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$

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



## Regularization parameters selection

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

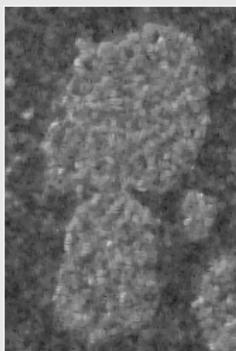
Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$

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



Co-localized contours estimate  $\hat{\mathbf{h}}^C$

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



too small

## Regularization parameters selection

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

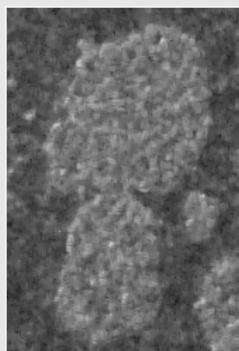
Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$

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



Co-localized contours estimate  $\hat{\mathbf{h}}^C$

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



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



too small

too large

## Regularization parameters selection

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

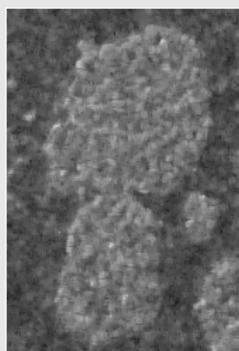
Lin. reg.  $\hat{\mathbf{h}}^{\text{LR}}$

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

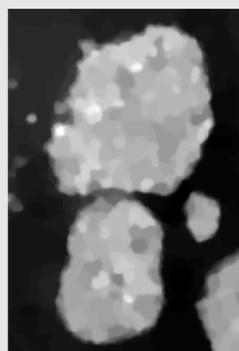


Co-localized contours estimate  $\hat{\mathbf{h}}^C$

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



$$(\lambda^\dagger, \alpha^\dagger) = (11.5, 0.8)$$



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



too small

optimal

too large

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

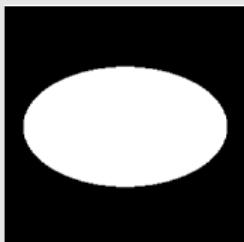
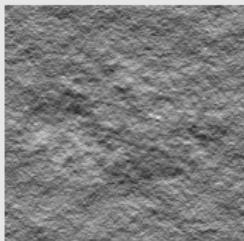
## Parameter tuning (Grid search)

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

## Parameter tuning (Grid search)

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

### Example



## Parameter tuning (Grid search)

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

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

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

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

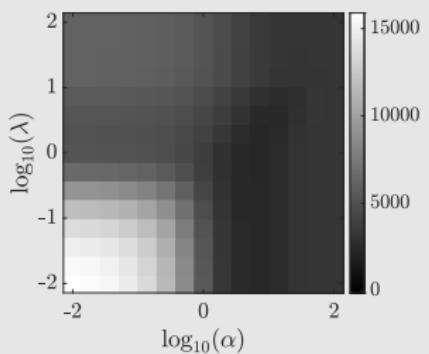
## Parameter tuning (Grid search)

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

$\mathbf{h}$ : discriminant,  $\mathbf{v}$ : auxiliary

$\bar{\mathbf{h}}$ : true regularity

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



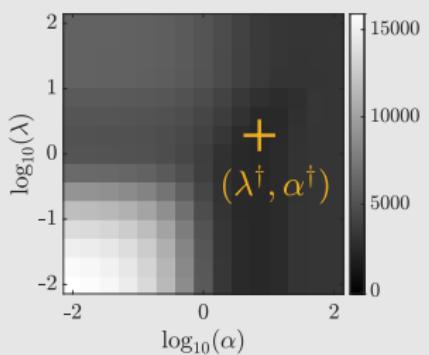
## Parameter tuning (Grid search)

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

$\mathbf{h}$ : discriminant,  $\mathbf{v}$ : auxiliary

$\bar{\mathbf{h}}$ : true regularity

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



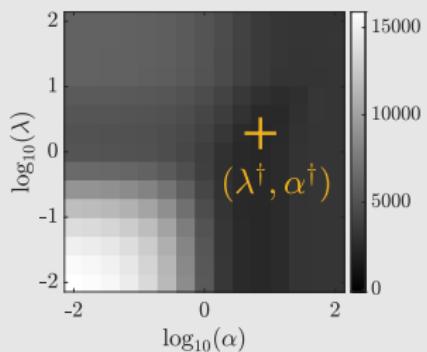
## Parameter tuning (Grid search)

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

$\mathbf{h}$ : discriminant,  $\mathbf{v}$ : auxiliary

$\bar{\mathbf{h}}$ : true regularity

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



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

?

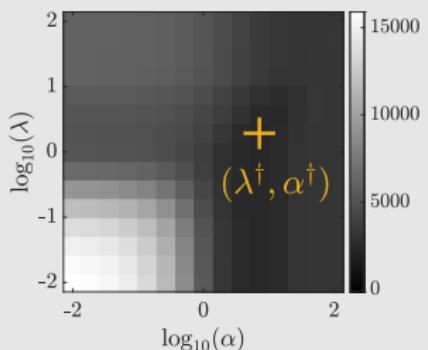
## Parameter tuning (Grid search)

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

$\mathbf{h}$ : discriminant,  $\mathbf{v}$ : auxiliary

$\bar{\mathbf{h}}$ : true regularity

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



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

?

Stein Unbiased Risk Estimate  
(SURE)

## *Stein Unbiased Risk Estimate (Principle)*

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

## *Stein Unbiased Risk Estimate* (Principe)

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

**Parametric estimator**  $(\mathbf{y}; \lambda) \mapsto \hat{\mathbf{x}}(\mathbf{y}; \lambda)$

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

## Stein Unbiased Risk Estimate (Principe)

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

**Parametric estimator**  $(\mathbf{y}; \lambda) \mapsto \hat{\mathbf{x}}(\mathbf{y}; \lambda)$

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

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

# Stein Unbiased Risk Estimate (Principe)

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

**Parametric estimator**  $(\mathbf{y}; \lambda) \mapsto \hat{\mathbf{x}}(\mathbf{y}; \lambda)$

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

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

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

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

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

$$\begin{aligned} \widehat{R}(\mathbf{y}; \lambda) &\triangleq \|\hat{\mathbf{x}}(\mathbf{y}; \lambda) - \mathbf{y}\|^2 + 2\rho^2 \operatorname{tr}(\partial_{\mathbf{y}} \hat{\mathbf{x}}(\mathbf{y}; \lambda)) - \rho^2 P \\ &\implies R(\lambda) = \mathbb{E}_{\boldsymbol{\zeta}} [\widehat{R}(\mathbf{y}; \lambda)]. \end{aligned}$$

## Generalized Stein Unbiased Risk Estimate

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

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

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

$$\Phi : (\mathbf{h}, \mathbf{v}) \mapsto \{\log(a)\mathbf{h} + \mathbf{v}\}_a \quad \begin{array}{|c|c|c|c|c|c|c|c|} \hline & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \Pi : (\mathbf{h}, \mathbf{v}) \mapsto (\mathbf{h}, \mathbf{0})$$

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

# Generalized Stein Unbiased Risk Estimate

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

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

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

$$\Phi : (\mathbf{h}, \mathbf{v}) \mapsto \{\log(a)\mathbf{h} + \mathbf{v}\}_a \quad \begin{array}{|c|c|c|c|c|c|c|c|c|c|c|c|} \hline & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \Pi : (\mathbf{h}, \mathbf{v}) \mapsto (\mathbf{h}, \mathbf{0})$$

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

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

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

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

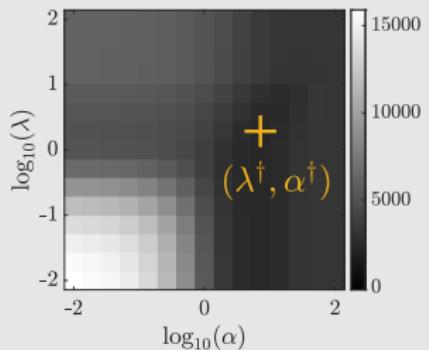
$$\begin{aligned} \hat{R}(\Lambda) &\triangleq \|\mathbf{A}(\Phi \hat{\mathbf{x}}(\mathbf{y}; \Lambda) - \mathbf{y})\|^2 + 2\text{tr} \left( \mathcal{S} \mathbf{A}^\top \Pi \partial_{\mathbf{y}} \hat{\mathbf{x}}(\mathbf{y}; \Lambda) \right) - \text{tr} \left( \mathbf{A} \mathcal{S} \mathbf{A}^\top \right) \\ &\implies R_{\Pi}(\Lambda) = \mathbb{E}_{\zeta} [\hat{R}(\Lambda)]. \end{aligned}$$

## Parameter tuning (Grid search)

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

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

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



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

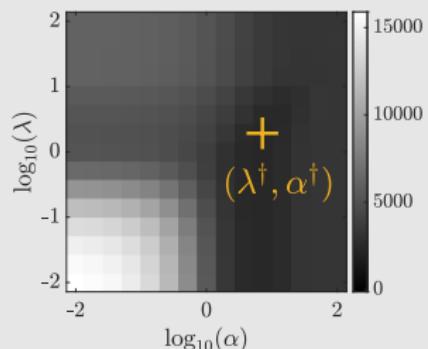
$$\widehat{\mathcal{R}}_{\nu, \epsilon}(\mathcal{L}; \lambda, \alpha | \mathcal{S})$$

## Parameter tuning (Grid search)

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

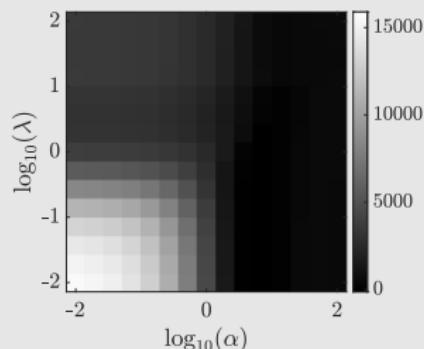
$\bar{\mathbf{h}}$ : true regularity

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



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

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

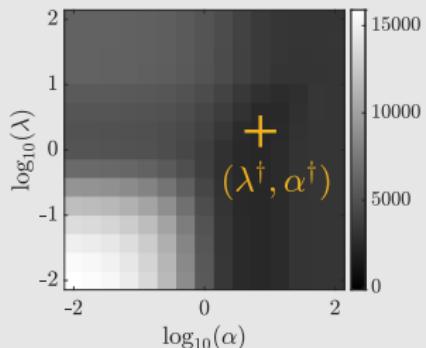


## Parameter tuning (Grid search)

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

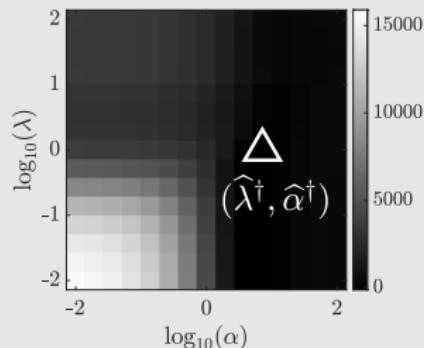
$\bar{\mathbf{h}}$ : true regularity

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



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

$$\hat{\mathcal{R}}_{\nu, \epsilon}(\mathcal{L}; \lambda, \alpha | \mathcal{S})$$

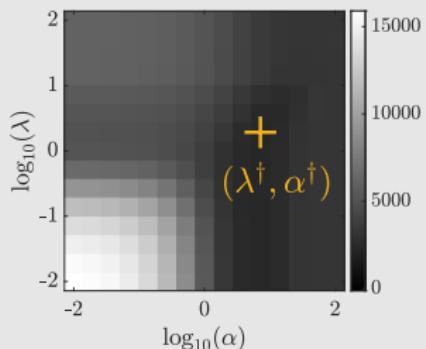


## Parameter tuning (Grid search)

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

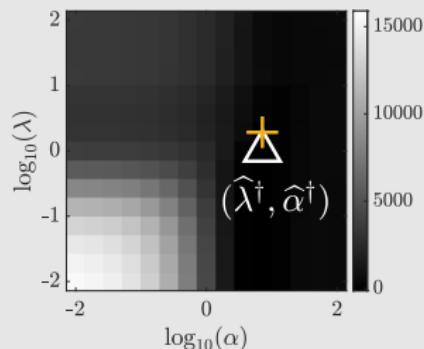
$\bar{\mathbf{h}}$ : true regularity

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



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

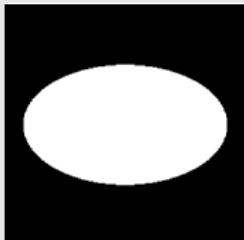
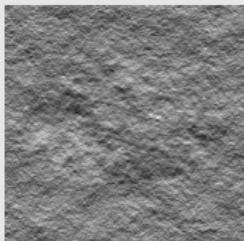
$$\hat{\mathcal{R}}_{\nu, \epsilon}(\mathcal{L}; \lambda, \alpha | \mathcal{S})$$



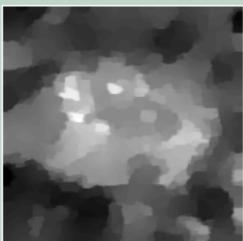
## Parameter tuning (Grid search)

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

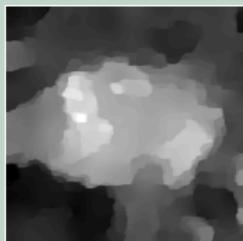
Example



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

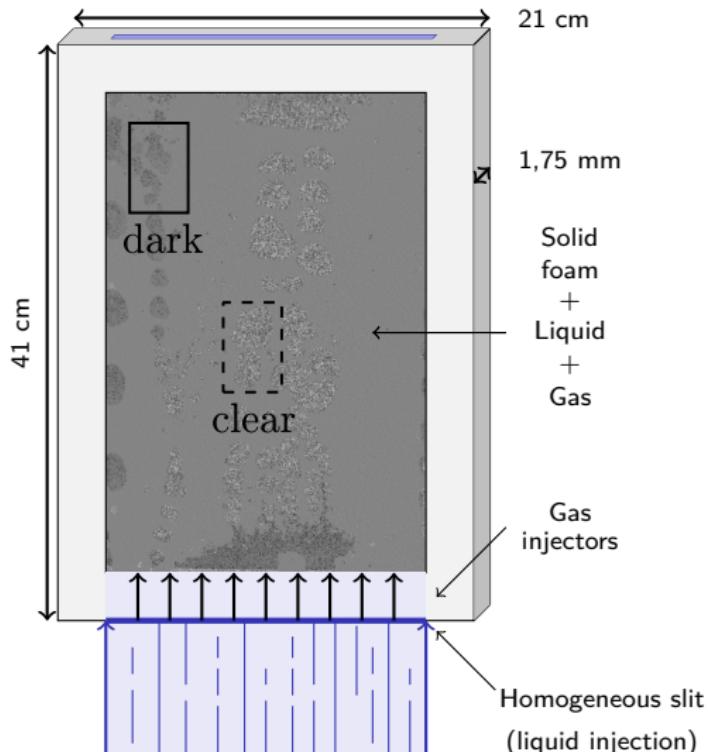


$\widehat{\mathbf{h}}^F(\mathcal{L}; \widehat{\lambda}^\dagger, \widehat{\alpha}^\dagger)$   
(grid)



# Multiphase flow through porous media

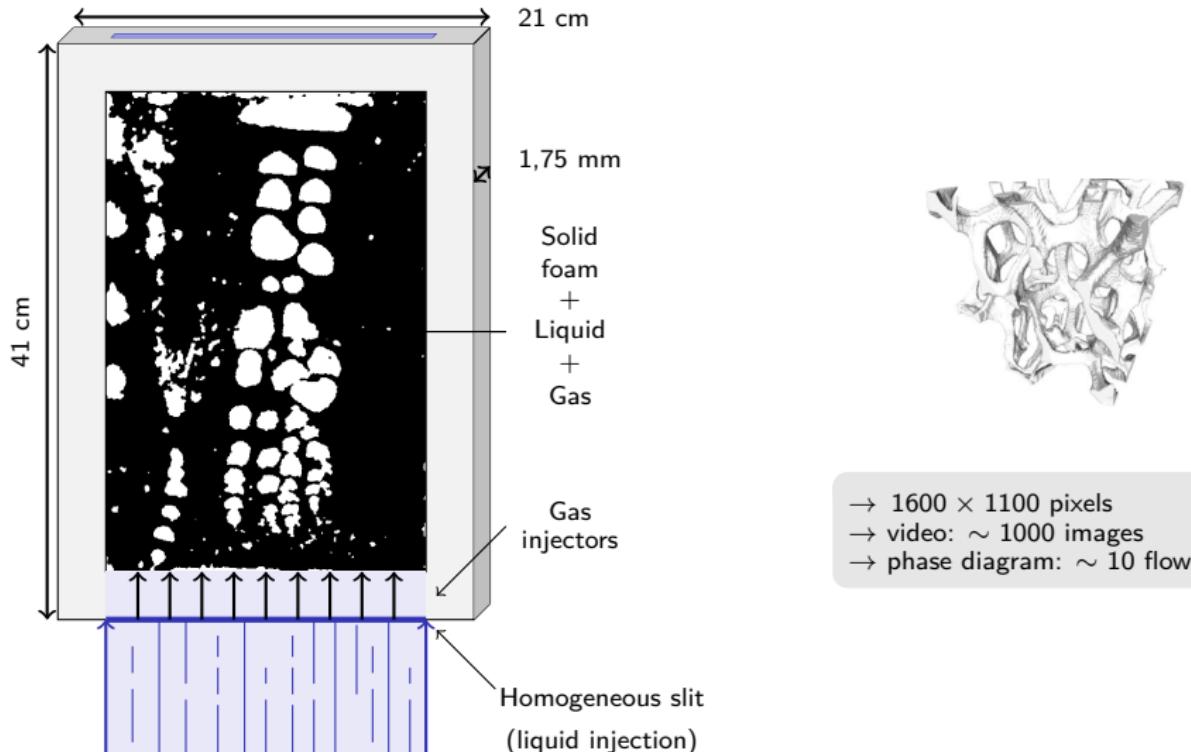
Laboratoire de Physique, ENS Lyon, V. Vidal, T. Busser, (M. Serres, IFPEN)



- 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)

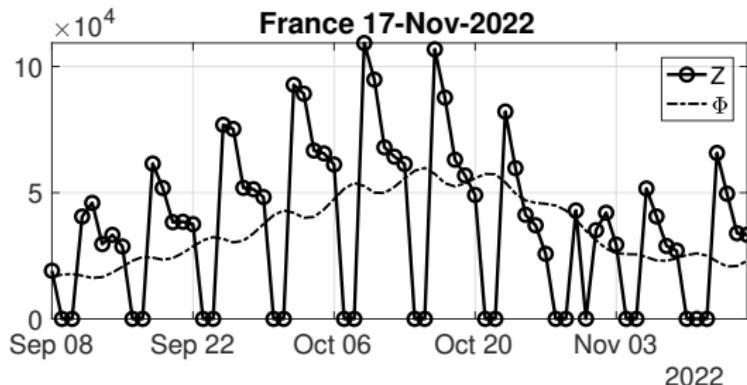


Time series analysis:

Epidemiological indicator estimation

# Epidemic propagation: modeling at the service of monitoring

## Counts of daily new infections

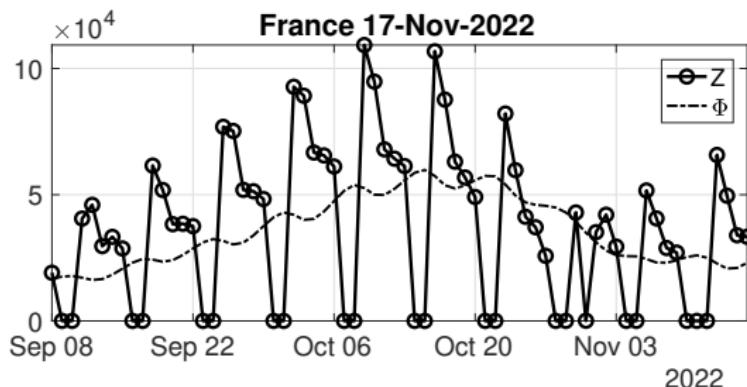


data from National Health Agencies collected by Johns Hopkins University

⇒ number of cases not informative enough: need to capture the **dynamics**

# Epidemic propagation: modeling at the service of monitoring

## Counts of daily new infections



data from National Health Agencies collected by Johns Hopkins University

⇒ number of cases not informative enough: need to capture the **dynamics**

Design adapted counter measures and evaluate their effectiveness

→ efficient monitoring tools

*epidemiological model,*

→ robust to low quality of the data

*managing erroneous counts.*

# Pandemic study: modeling at the service of monitoring

## 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*)

# Pandemic study: modeling at the service of monitoring

## 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.

⇒ one single indicator accounting for the overall pandemic mechanism

# Pandemic study: modeling at the service of monitoring

## 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.

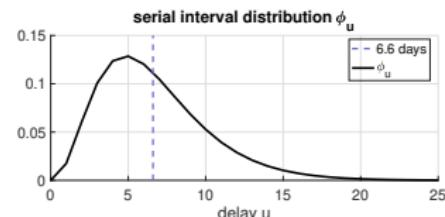
⇒ one single indicator accounting for the overall pandemic mechanism

**Principle:**  $Z_t$  new infections at day  $t$

$$\mathbb{E}[Z_t] = R_t \Phi_t, \quad \Phi_t = \sum_{u=1}^{\tau_\Phi} \phi_u Z_{t-u}$$

with  $\Phi_t$  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

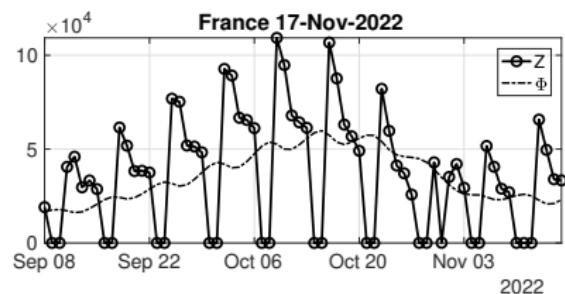
# Pandemic study: modeling at the service of monitoring

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

**Model:** **scaled** Poisson distribution

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

$\gamma > 0$  scaling parameter: controls **variance**  
(Pascal & Vaiter, 2025, *Signal Process.*)



# Pandemic study: modeling at the service of monitoring

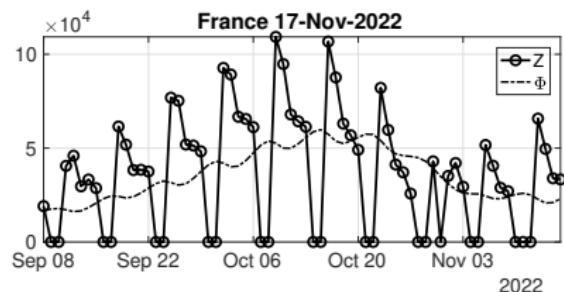
**Data:** daily counts  $\mathbf{Z} = (Z_1, \dots, Z_T)$

**Model:** scaled Poisson distribution

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

$\gamma > 0$  scaling parameter: controls variance

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



**Inverse problem formalism:**

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

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

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

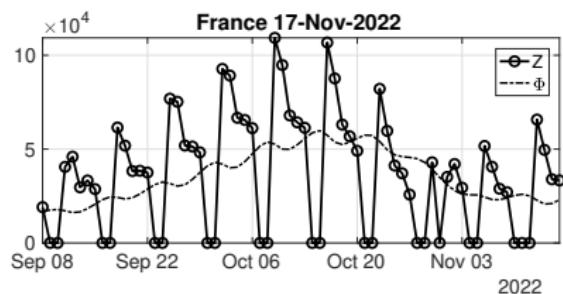
# Pandemic study: modeling at the service of monitoring

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

**Model:** scaled Poisson distribution

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

$\gamma > 0$  scaling parameter: controls variance  
(Pascal & Vaiter, 2025, *Signal Process.*)



## Maximum Likelihood Estimate (MLE)

$$\begin{aligned} & \ln (\mathbb{P}(Z_t | \mathbf{Z}_{t-\tau_\Phi:t-1}, R_t)) \\ &= \underset{\text{(def.)}}{-\frac{1}{\gamma} d_{KL}(Z_t | R_t \Phi_t)} \quad \text{(Kullback-Leibler)} \end{aligned}$$

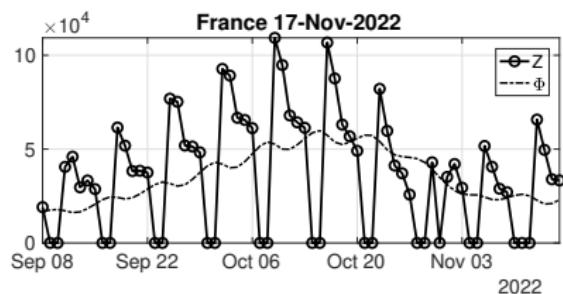
# Pandemic study: modeling at the service of monitoring

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

**Model:** scaled Poisson distribution

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

$\gamma > 0$  scaling parameter: controls variance  
(Pascal & Vaiter, 2025, *Signal Process.*)



## Maximum Likelihood Estimate (MLE)

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

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

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

ratio of moving averages

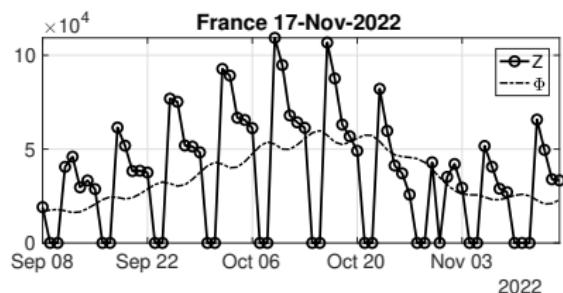
# Pandemic study: modeling at the service of monitoring

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

**Model:** scaled Poisson distribution

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

$\gamma > 0$  scaling parameter: controls variance  
(Pascal & Vaiter, 2025, *Signal Process.*)



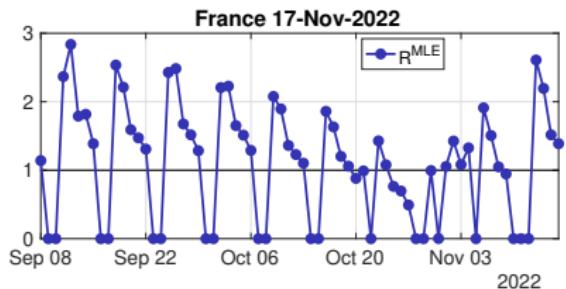
## Maximum Likelihood Estimate (MLE)

$$\ln(\mathbb{P}(Z_t | Z_{t-\tau_\Phi:t-1}, R_t))$$

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

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

ratio of moving averages



- huge variability along time/ no local trend
- not robust to pseudo-periodicity/ misreported counts

Penalized likelihood: regularization through nonlinear filtering

$$\widehat{\mathbf{R}}^{\text{PKL}} = \underset{\mathbf{R} \in \mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{\gamma} d_{\text{KL}}(Z_t | R_t \Phi_t) + \lambda \mathcal{R}(\mathbf{R}) \quad (\text{penalized Kullback-Leibler})$$

with  $\mathcal{R}(\mathbf{R})$  favoring some temporal regularity

(Abry et al., 2020, *PLOS One*)

Penalized likelihood: regularization through nonlinear filtering

$$\hat{\mathbf{R}}^{\text{PKL}} = \underset{\mathbf{R} \in \mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{\gamma} d_{\text{KL}}(Z_t | R_t \Phi_t) + \lambda \mathcal{R}(\mathbf{R}) \quad (\text{penalized Kullback-Leibler})$$

with  $\mathcal{R}(\mathbf{R})$  favoring some temporal regularity

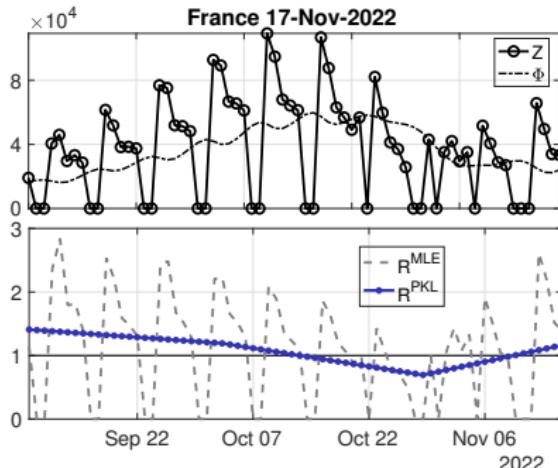
(Abry et al., 2020, *PLOS One*)

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

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

2nd order derivative &  $\ell_1$ -norm

⇒ piecewise linearity



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

Penalized Kullback-Leibler estimator:

$$\widehat{\mathbf{R}}(\mathbf{Z}; \lambda) = \underset{\mathbf{R} \in \mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{\gamma} d_{KL} (\mathbf{Z}_t | \mathbf{R}_t \boldsymbol{\Phi}_t) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1$$

with  $\mathcal{R}(\mathbf{R})$  favoring some temporal regularity

(Abry et al., 2020, *PLOS One*)

## Penalized Kullback-Leibler estimator:

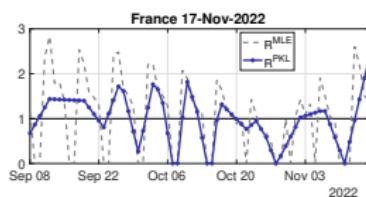
$$\widehat{\mathbf{R}}(\mathbf{Z}; \lambda) = \underset{\mathbf{R} \in \mathbb{R}_+^T}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{\gamma} d_{KL} (\mathbf{Z}_t | \mathbf{R}_t \boldsymbol{\Phi}_t) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1$$

with  $\mathcal{R}(\mathbf{R})$  favoring some temporal regularity

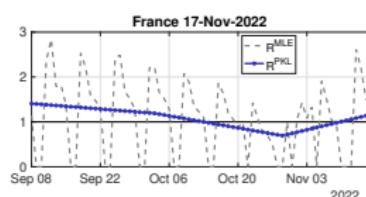
(Abry et al., 2020, *PLOS One*)

## Fine tuning of the regularization parameter:

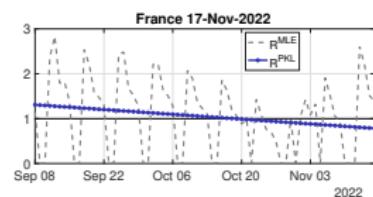
$$\lambda = 3.5$$



$$\lambda^\dagger = 50$$



$$\lambda = 250$$



## Penalized Kullback-Leibler estimator:

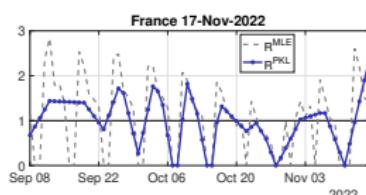
$$\widehat{\mathbf{R}}(\mathbf{Z}; \lambda) = \operatorname{argmin}_{\mathbf{R} \in \mathbb{R}_+^T} \sum_{t=1}^T \frac{1}{\gamma} d_{KL} (\mathbf{Z}_t | \mathbf{R}_t \Phi_t) + \lambda \|\mathbf{D}_2 \mathbf{R}\|_1$$

with  $\mathcal{R}(\mathbf{R})$  favoring some temporal regularity

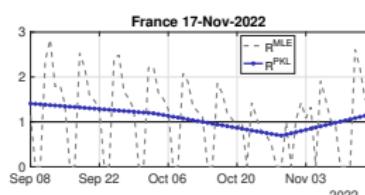
(Abry et al., 2020, *PLOS One*)

## Fine tuning of the regularization parameter:

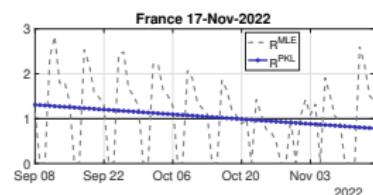
$$\lambda = 3.5$$



$$\lambda^\dagger = 50$$



$$\lambda = 250$$



## Data-driven oracle minimization

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

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

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

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

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

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

**Reminder:**

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

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

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

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

**Reminder:**

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

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

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

**Reminder:**

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

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

**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:  $(\Phi\mathbf{R})_t = R_t \sum_{s=1}^{\tau_\Phi} \phi_s \mathbf{Z}_{t-s}$   
 $\implies$  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  $\mathbb{E}_{\mathbf{Z}} [\mathcal{O}(\mathbf{Z}; \lambda)] = \mathbb{E}_{\mathbf{Z}} \left[ \|\widehat{\mathbf{R}}(\mathbf{Z}; \lambda) - \overline{\mathbf{R}}\|_2^2 \right]$

**Reminder:**

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

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

**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:  $(\Phi\mathbf{R})_t = R_t \sum_{s=1}^{\tau_\Phi} \phi_s \mathbf{Z}_{t-s}$   
 $\implies$  Novel counterpart of Stein lemma for driven autoregressive Poisson model

Autoregressive Poisson Unbiased Risk Estimate (APURE)

## Model

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

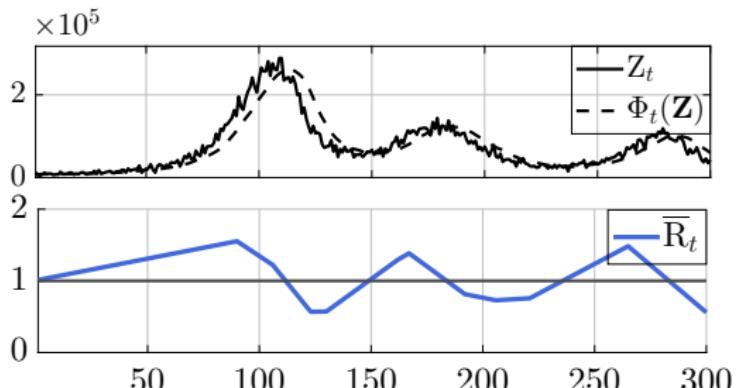
with scale parameter  $\log_{10} \gamma \equiv 3$

# Data-driven hyperparameter selection on synthetic Poisson data

## Model

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

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



# Data-driven hyperparameter selection on synthetic Poisson data

## Model

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

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

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

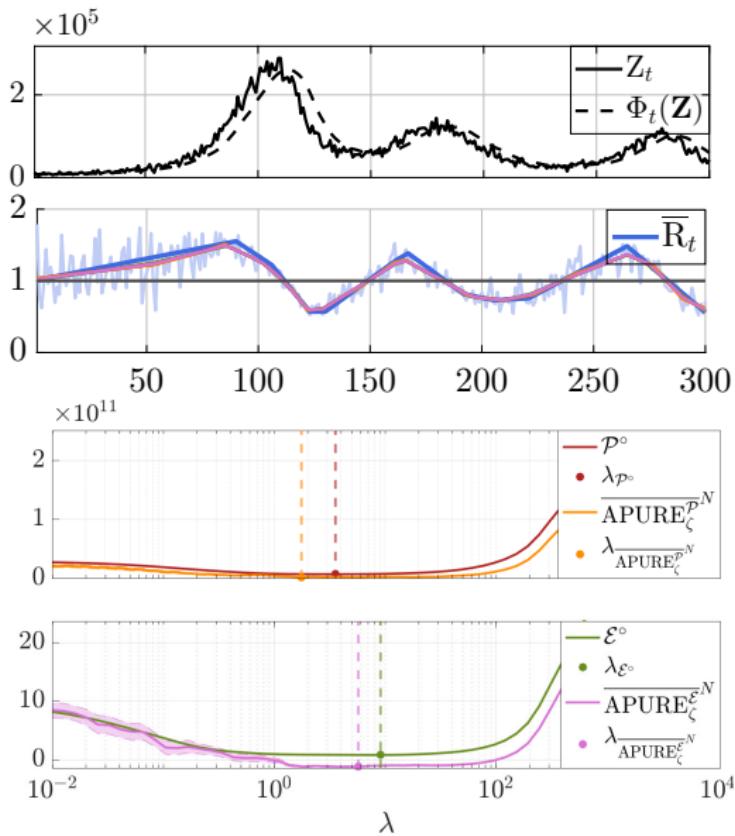
$$\mathcal{P}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) \odot \Phi(Z) - \bar{R} \odot \Phi(Z)\|_2^2$$

$$\mathcal{E}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) - \bar{R}\|_2^2$$

## Unbiased risk estimators

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

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



# Data-driven hyperparameter selection on synthetic Poisson data

## Model

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

with scale parameter  $\log_{10} \gamma \equiv 3.5$

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

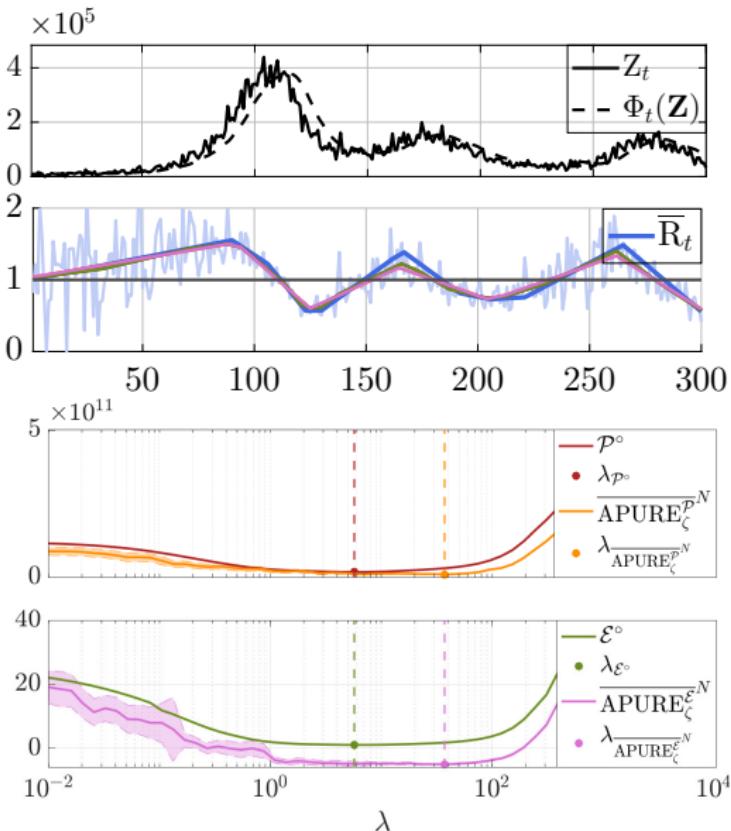
$$\mathcal{P}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) \odot \Phi(Z) - \bar{R} \odot \Phi(Z)\|_2^2$$

$$\mathcal{E}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) - \bar{R}\|_2^2$$

## Unbiased risk estimators

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

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



# Data-driven hyperparameter selection on synthetic Poisson data

## Model

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

with scale parameter  $\log_{10} \gamma \equiv 4$

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

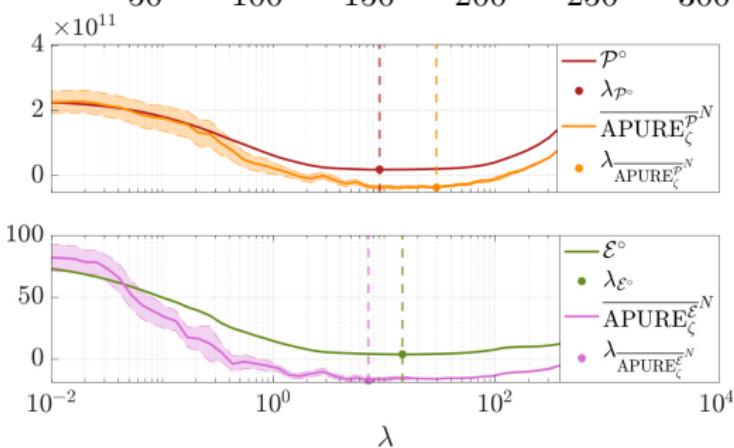
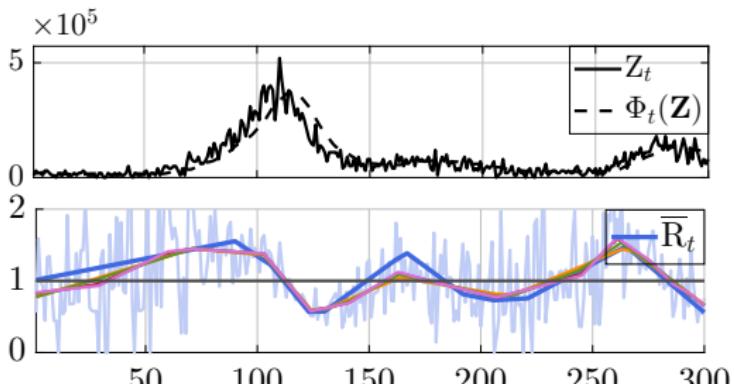
$$\mathcal{P}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) \odot \Phi(Z) - \bar{R} \odot \Phi(Z)\|_2^2$$

$$\mathcal{E}^\circ(Z; \theta) = \|\hat{R}(Z; \theta) - \bar{R}\|_2^2$$

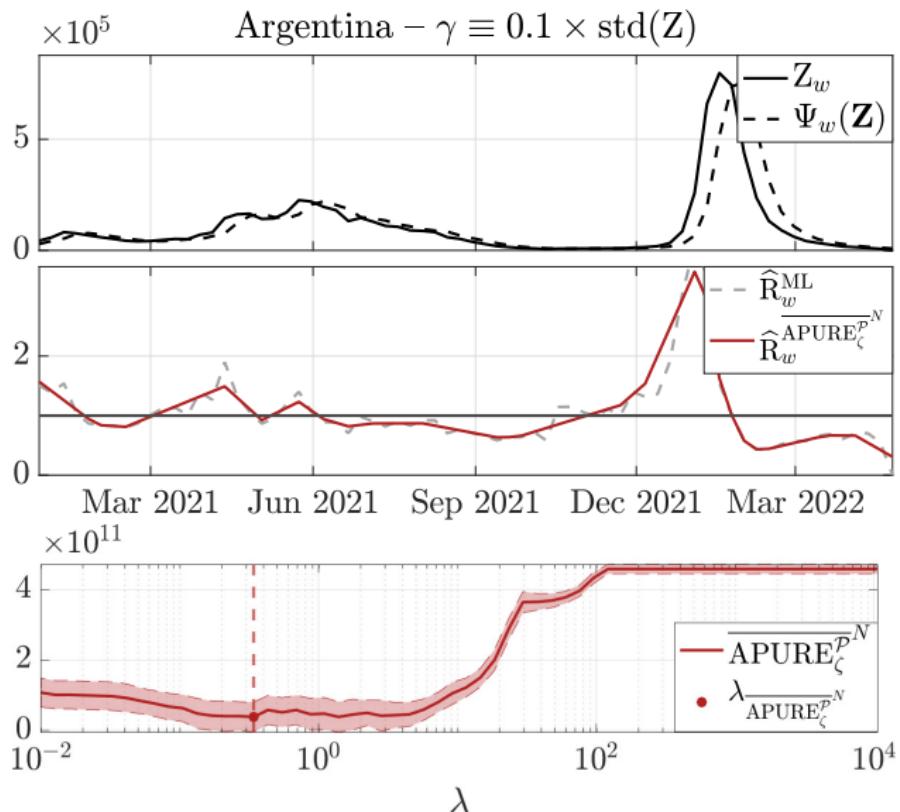
## Unbiased risk estimators

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

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



# Data-driven hyperparameter selection on weekly COVID-19 counts



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

Codes: [github.com/bpascal-fr/APURE-Estim-Epi](https://github.com/bpascal-fr/APURE-Estim-Epi)

## Conclusion and perspectives

**Inverse problem**

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

$$\lambda^\dagger \in \operatorname*{Argmin}_{\lambda \in \Lambda} \mathcal{O}(\mathbf{y}; \lambda), \quad \text{for} \quad \hat{\mathbf{x}}(\mathbf{y}; \lambda) \in \operatorname*{Argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{D}(\mathbf{y}, \Phi \mathbf{x}) + \lambda \mathcal{R}(\mathbf{x})$$

## Conclusion and perspectives

### Inverse problem

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

$$\lambda^\dagger \in \operatorname*{Argmin}_{\lambda \in \Lambda} \mathcal{O}(\mathbf{y}; \lambda), \quad \text{for} \quad \hat{\mathbf{x}}(\mathbf{y}; \lambda) \in \operatorname*{Argmin}_{\mathbf{x} \in \mathbb{R}^N} \mathcal{D}(\mathbf{y}, \Phi \mathbf{x}) + \lambda \mathcal{R}(\mathbf{x})$$

### Data-driven parameter selection

⇒  $\mathcal{O}$ : 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.

# Conclusion and perspectives

## Inverse problem

$$\mathbf{y} \sim \mathcal{B}(\Phi \bar{\mathbf{x}})$$

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

## Data-driven parameter selection

⇒  $\mathcal{O}$ : 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  $\mathcal{R}(\mathbf{x})$ ;
- ▶ Generalization to other noise models: speckle noise in medical imaging;
- ▶ Unsupervised learning for  $\hat{\mathbf{x}}(\mathbf{y}; \lambda) = \mathbf{NN}_\theta(\mathbf{y})$  with loss  $\mathcal{O}(\mathbf{y}; \theta)$ .