Plan	Microarray data analysis	lssue	Some solutions 00000 000 000000000	New framework 0000000 000000000 00000000 0000000	
Manifolds, a new framework for warped curves analysis Séminaires de l'unité BIA (INRA/MIA Toulouse)					
		Elie Maz	A		
INP-ENSAT – INRA (laboratoire GBF)					

9 décembre 2011

Elie Maza

-INP-ENSAT - INRA (laboratoire GBF)

A = A + A = A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 0000 00000000	New framework 0000000 000000000 00000000 0000000 0000
ΡI	an			
1 2	Microarray data analysis Issue			
3	Some solutions <ul> <li>Shift model</li> <li>Phase model</li> <li>Quantile normalization</li> </ul>			
4	New framework			

INP-ENSAT - INRA (laboratoire GBF)

- Idea
- $\bullet$  Approximation of  $\delta$
- Shift model
- Application

#### Elie Maza

Plan	Microarray data analysis	Issue	Some solutions 00000 000 000000000	New framework 0000000 000000000 00000000 0000000
Plan				
1 Mic	roarray data analysis			
<ol> <li>Issue</li> </ol>				
3 Som • Sl • P • Q	ne solutions hift model hase model Quantile normalization			
<ul><li>4 New</li><li>Id</li></ul>	/ framework lea			
• A	pproximation of $\delta$			
• SI	hift model			
• A	pplication			
				トイヨト ヨークへ(~ NDA (laborateire CDE)
Elle Maza			INP-ENSAL – I	NRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000

Central Dogma of Molecular Biology (F. CRICK, 1970, Nature)



Fig. 3. A tentative classification for the present day. Solid arrows show general transfers; dotted arrows show special transfers. Again, the absent arrows are the undetected transfers specified by the central dogma.

→ The study of the genetic information contained in any organism:

- Finding coding sequences in the DNA.
- Measuring the abundance of RNAs.
- Studing the diversity of Proteins.

INP-ENSAT - INRA (laboratoire GBF)

Manifolds, a new framework for warped curves analysis

Elie Maza



DNA Microarray : A "new" technology for transcriptome studies.



National Human Genome Research Institute, http://www.genome.gov/

#### Elie Maza

INP-ENSAT - INRA (laboratoire GBF)



The ratio of red and green fluorescence intensities for each spot is intended to be indicative of the relative abundance of the transcripts in the test condition compared to the reference condition:

$$M = \log_2\left(\frac{R}{G}\right) = \log_2(R) - \log_2(G)$$
$$A = \frac{1}{2}\left(\log_2(R) + \log_2(G)\right) = \log_2\left(\sqrt{RG}\right)$$

INP-ENSAT - INRA (laboratoire GBF)

(日) (同) (三) (

Manifolds, a new framework for warped curves analysis

Elie Maza

Plan	Microarray data analysis	Issue	Some solutions 00000 000 000000000	New framework 0000000 000000000 00000000 0000000
Example				

- Genomic and Biotechnology of the Fruit laboratory (GBF)
- tomato microarrays
- Data:
  - 18 conditions
  - $13056 \times 2$  measures per microarray

Issue

Some solutions

#### New framework

0000000 000000000 00000000 0000000

And         And <th>2-range 5.1 to 6.1 (attemption 5.1, 6.1)</th> <th>2-angs 5.7 to 6.6 (saturation 5.7, 6.6)</th> <th>r-range 5.1 to 6.3 (saluation 5.1, 6.3)</th>	2-range 5.1 to 6.1 (attemption 5.1, 6.1)	2-angs 5.7 to 6.6 (saturation 5.7, 6.6)	r-range 5.1 to 6.3 (saluation 5.1, 6.3)
	No.         No. <td>2 - mrg 6 2 to 7.4 (saturation 6.2, 7.4)</td> <td>2-range 55 to 69 (subardion 55, 6.9)</td>	2 - mrg 6 2 to 7.4 (saturation 6.2, 7.4)	2-range 55 to 69 (subardion 55, 6.9)
2-range 5.6 to E.1 (surantion 5.6, 8.1)	2-range 52 to 6.5 (mitamation 52, 6.5)	2-mage 6.7 to 7.2 (statution 5.7, 7.2)	
2-range 5.1 to 8.4 (asturation 5.1, 8.4)	2010 100 100 100 100 100 100 100 100 100	2	2 Construction 4.3 (C) 2 Construction 4.3 (C) 2 Construction 4.3 (C)
Sect. Solid         Sect. Solid	Mail         Mail <th< td=""><td>2-ange 5.8 to 8.4 (parentice 5.8, 8.4)</td><td>z-tangat 5.1 to 7 (naturation 5.1,7)</td></th<>	2-ange 5.8 to 8.4 (parentice 5.8, 8.4)	z-tangat 5.1 to 7 (naturation 5.1,7)
200 100 100 100 100 100 100 100 100 100	2-range 5.1 to 7.9 (out-railion 5.1, 7.9)	2-ange 6 to 8 (usualistic 6, 8)	2-range 52 to 55 (naturation 52, 65)
2-carge 52 to 7 (astunitor 52, 7)	2-range 48 to 57 (saturation 48, 57)	2-encys 5.3 to 7.1 (saturation 5.3, 7.1)	100         100
State         State <th< td=""><td>r-range 4.9 to 7.6 (saturation 4.9, 7.6)</td><td>2-mage 5.6 to £.1 (astronico 5.6, £.1)</td><td>2-range 52 to 56 (tailuration 52,96)</td></th<>	r-range 4.9 to 7.6 (saturation 4.9, 7.6)	2-mage 5.6 to £.1 (astronico 5.6, £.1)	2-range 52 to 56 (tailuration 52,96)

Elie Maza

Manifolds, a new framework for warped curves analysis

INP-ENSAT - INRA (laboratoire GBF)

2

・ロト ・聞ト ・ヨト ・ヨト

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000

High variability of microarray data:

- RNA extraction
- hybridization conditions (temperature, humidity, ...)
- image acquisition
- heat and light sensitivities for Cy3 (Green) and Cy5 (Red)

• . . .

→ Normalization procedures!

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000

Two "standard" normalization procedures:

- within arrays (print-tip loess)
- between arrays (quantile normalization)

Software package limma:

 Smyth, G. K. (2005). Limma: linear models for microarray data. In: 'Bioinformatics and Computational Biology Solutions using R and Bioconductor'. R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, W. Huber (eds), Springer, New York, pages 397–420.

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



INP-ENSAT – INRA (laboratoire GBF)



INP-ENSAT - INRA (laboratoire GBF)



INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions 00000 000000000	New framework 0000000 000000000 00000000 0000000
Plan				
<ol> <li>Micro</li> <li>Issue</li> <li>Some         <ul> <li>Shi</li> <li>Pha</li> <li>Qui</li> </ul> </li> </ol>	parray data analysis solutions ft model ase model antile normalization			
<ul> <li>A New 1</li> <li>Ide</li> <li>App</li> <li>Shi</li> <li>App</li> </ul>	framework a proximation of $\delta$ ft model plication		< D > < 4 > < 3	> < ≅ > ∃ ∽00 @
Elie Maza			INP-ENSAT – I	NRA (laboratoire GBF)
Manifolds a n	w framework for warned curves analy	vsis		

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000

Warping model:

$$G = F \circ H^{-1}$$

where

• F is a stochastic function such as  $\mathbb{E}(F) = f$  (amplitude variation)

• *H* is a strictly increasing stochastic function such as  $\mathbb{E}(H) = \phi$  (phase variation)

The structural expectation:

$$f_{\rm ES} \stackrel{
m def}{=} f \circ \phi^{-1}$$

Elie Maza

Manifolds, a new framework for warped curves analysis

INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000

## Data:

$$Y_{ij} = F_i \circ H_i^{-1}(t_{ij}) + \epsilon_{ij}, \ i = 1, \dots, n, \ j = 1, \dots, m$$

where

- $F_i \sim F$  are *iid*
- $H_i \sim H$  are *iid*
- $t_{ij} \in [a, b] \subset \mathbb{R}$
- $\epsilon_{ij}$  are *iid* with mean 0 and variance  $\sigma^2$

(日) (同) (三) (

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



#### INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000



#### INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 000 000000000	New framework 0000000 00000000 00000000 0000000

# Solution:

- Alignement of curves.
- 2 Estimation of mean of aligned curves.

Elie Maza

INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New fra 00000 00000 00000 00000	amework 00 0000 000 0
Plan					
<ol> <li>Micro</li> <li>Issue</li> <li>Some         <ul> <li>Some</li> <li>Sh</li> <li>Ph</li> <li>Qu</li> </ul> </li> </ol>	oarray data analysis e solutions ift model nase model uantile normalization				
<ul> <li>4 New</li> <li>• Ide</li> <li>• Appendix</li> <li>• Sh</li> <li>• Appendix</li> </ul>	framework ea oproximation of $\delta$ ift model oplication		< <b>□</b> > < <b>∂</b> > < ≅	> 《동 > 문	৶৶৻৻
Elie Maza			INP-ENSAT – II	NRA (laboratoi	re GBF)

Plan	Microarray data analysis	Issue	Some solutions ●0000 ○○○ ○○○○○○○○○○	New framework 0000000 000000000 00000000 00000000
Shift mode	I			
Plan				
1 Mic	croarray data analysis			
2 Issu				
3 Sor • S • F • C	<b>ne solutions</b> Shift model Phase model Quantile normalization			
<ul> <li>4 Nev</li> <li>6 A</li> <li>6 A</li> <li>7 A</li> </ul>	w framework dea Approximation of $\delta$ Shift model Application			
			<□> <@> <≡	<ul><li>&lt; ≣ &lt; &lt;</li></ul>
Elie Maza			INP-ENSAT – II	NRA (laboratoire GBF)
Manifolds,	a new framework for warped curves analy	sis		

Plan	Microarray data analysis	Issue	<b>Some solutions</b> ○●○○○ ○○○○○○○○○○	New framework 0000000 00000000 00000000 0000000 00000
Shift model				

# Model:

$$Y_{ij} = f(t_j - \theta_i^*) + \epsilon_{ij}, i = 1, \dots, n, j = 1, \dots, m,$$

with

Elie Maza

f: ℝ → ℝ an unknown *T*-periodic function
θ = (θ<sub>i</sub><sup>\*</sup>)<sub>i=1,...,n</sub> an unknown shift parameter
∀j = 1,..., m, t<sub>j</sub> = (j-1)/m T ⊂ [0, T[
∀i = 1,..., n, (ε<sub>ij</sub>)<sub>j=1,...,m</sub> are *iid* and N(0, 1)

INP-ENSAT - INRA (laboratoire GBF)

(日) (同) (三) (

Plan	Microarray data analysis	Issue	Some solutions ○○●○○ ○○○○○○○○○○	New framework 0000000 00000000 0000000 0000000 000000
Shift model				

Other model:

$$d_{il} = e^{-il\alpha_i^*}c_l(f) + w_{il}, \ i = 1, \dots, n, \ l = -(m-1)/2, \dots, (m-1)/2,$$

with

•  $(c_l(f))_{l \in \mathbb{Z}}$  the Fourier coefficients of f:

$$\forall l \in \mathbb{Z}, \ c_l(f) = rac{1}{T} \int_0^T f(t) e^{-i2\pi rac{tl}{T}} \mathrm{d}t$$

α<sup>\*</sup> = (α<sup>\*</sup><sub>i</sub>)<sub>i=1,...,n</sub> = (<sup>2π</sup>/<sub>T</sub> θ<sup>\*</sup><sub>i</sub>)<sub>i=1,...,n</sub> the normalized shift parameter
 ∀i ∈ {1,...,n}, (w<sub>il</sub>)<sub>l=-(m-1)/2,...,(m-1)/2</sub> are complex *iid* variables with mean 0 and variance <sup>1</sup>/<sub>n</sub>

#### Elie Maza

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Shift model				

# Contraste:

$$M_n(\alpha) = \frac{1}{n} \sum_{i=1}^n \sum_{l=-\frac{m-1}{2}}^{l=\frac{m-1}{2}} \delta_l^2 \, |\tilde{c}_{il}(\alpha) - \hat{c}_l(\alpha)|^2$$

# with

c
<sub>il</sub>(α) = e<sup>ilα<sub>i</sub></sup>d<sub>il</sub> are the scaled Fourier coefficients
 c
<sub>l</sub>(α) = <sup>1</sup>/<sub>n</sub> Σ
<sub>i=1</sub><sup>n</sup> c
<sub>il</sub>(α) are the mean coefficients
 (δ<sub>l</sub>)<sub>l∈Z</sub> is such that Σ<sub>l∈Z</sub> δ
<sub>l</sub><sup>2</sup> < +∞</li>

Remark:  $\tilde{c}_{il}(\alpha^*) = c_l(f) + e^{il\alpha_i^*} w_{il}$  and  $\hat{c}_l(\alpha^*) = c_l(f) + \frac{1}{n} \sum_{i=1}^n e^{il\alpha_i^*} w_{il}$ 

Elie Maza

Manifolds, a new framework for warped curves analysis

INP-ENSAT - INRA (laboratoire GBF)

イロト イポト イヨト イヨト

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 0000● 0000000000000000000000000000000	New framework 0000000 00000000 0000000 0000000 000000
Shift model				

# **Semi-parametric Estimation of Shifts.** F. GAMBOA, J-M. LOUBES AND E. MAZA. *Electronic Journal of Statistics*, Vol. 1, 2007, 616–640.

Elie Maza

INP-ENSAT - INRA (laboratoire GBF)

Flati	WICFOAFFAY UALA ANAIYSIS	issue	00000000000000000000000000000000000000	0000000 00000000 00000000 0000000
Phase model				
Plan				
<ol> <li>Microa</li> <li>Issue</li> <li>Some         <ul> <li>Shif</li> <li>Pha</li> <li>Qua</li> </ul> </li> </ol>	array data analysis solutions t model se model antile normalization			
<ul> <li>4 New fr</li> <li>Idea</li> <li>App</li> <li>Shif</li> <li>App</li> </ul>	ramework proximation of $\delta$ it model lication		(0) (8) (3	> < ≅ > ≅ - ସେ∢୍ତ
Elie Maza			INP-ENSAT – I	NRA (laboratoire GBF)
Manifolds, a nev	v framework for warped curves analy	rsis		

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000000	00000000 00000000 0000000
Phase model				

## Model:

$$Y_{ij} = f \circ H_i^{-1}(t_j), \ i = 1, \dots, n, \ j = 1, \dots, m,$$

with

• 
$$f : [a, b] \rightarrow \mathbb{R}$$
 a continuous function

- $\forall j = 1, \ldots, m, t_j = a + (j-1) \frac{b-a}{m-1} \subset [a, b]$
- $H_i \sim H$  are *iid*, continuous and strictly increasing

A B A B A B A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A
 A



Idea: if f is strictly increasing, then

$$\left(f\circ H_i^{-1}\right)^{-1}=H_i\circ f^{-1}$$

hence

$$\mathbb{E}\left(\left(f\circ H_{i}^{-1}\right)^{-1}\right)=\mathbb{E}\left(H_{i}\right)\circ f^{-1}=\phi\circ f^{-1}=f_{\mathrm{ES}}^{-1}$$

Elie Maza

INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions ○○○○ ●○○○○○○○	<b>New framework</b> 0000000 000000000 00000000 0000000
Quantile r	normalization			
Plan				
1 Mi	croarray data analysis			
2 lss				
3 So • 1 •	me solutions Shift model Phase model Quantile normalization			
<ul> <li>A Ne</li> <li>A</li> <li>A<th>tw framework Idea Approximation of <math>\delta</math> Shift model Application</th><th></th><th></th><th></th></li></ul>	tw framework Idea Approximation of $\delta$ Shift model Application			
			▲□> ▲檀> ▲厘)	<ul><li>&lt; ≣ &lt; &lt; &lt; &gt; &lt; &lt;</li></ul>
Elie Maza			INP-ENSAT – II	NRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	<b>Some solutions</b> ○○○○ ○●○○○○○○○	New framework 0000000 000000000 00000000 000000
Quantile no	rmalization			
Bolsta	ad et al (2003)			

- Let  $X \sim F$  a random variable.
- Let  $X^1, \ldots, X^i, \ldots, X^m$  iid such that  $F_i = F$ .
- Let, for all  $i \in \{1, \ldots, m\}$ ,  $X_1^i, \ldots, X_j^i, \ldots, X_n^i$  iid such that  $X_j^i \sim X^i$ .
- Let, for all  $i \in \{1, \ldots, m\}$ ,  $X^i_{(1)}, \ldots, X^i_{(j)}, \ldots, X^i_{(n)}$  the order statistics.

The quantile normalization is defined, for all  $j \in \{1, \ldots, n\}$ , by

$$\hat{X}_{(j)} = \frac{1}{m} \sum_{i=1}^{m} X^{i}_{(j)}$$

(日) (同) (三) (

Elie Maza

Plan	Microarray data analysis	Issue	Some solutions ○○○○ ○○○ ○○●○○○○○○	New framework 0000000 000000000 0000000 0000000 00000
Quantile norm	alization			
Remarl	K			

For all  $i \in \{1, \ldots, m\}$ , assuming

$$X^{i}=H_{i}\left(X\right)$$

we have

$$F_i = F \circ H_i^{-1}$$

Remark: Bolstad et al. (2003) assume that H = Id.

Elie Maza

Manifolds, a new framework for warped curves analysis

INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions ○○○○○ ○○○●○○○○○○	New framework 0000000 00000000 00000000 0000000 00000
Quantile no	rmalization			
Proof				

 Non parametric estimation of the structural expectation of a stochastic increasing function. J.-F. DUPUY, J.-M. LOUBES and E. MAZA, Statistics and Computing, 2011.

$$\hat{F}^{-1} = \frac{1}{m} \sum_{i=1}^{m} \hat{F}_i^{-1} \xrightarrow{\mathbf{P}} F^{-1}$$

Statistical properties of the quantile normalization method for DNA microarray analysis. S. GALLÓN, J.-M. LOUBES and E. MAZA, preprint.

$$\hat{X}_{(j)} = \frac{1}{m} \sum_{i=1}^{m} X^{i}_{(j)} \xrightarrow{\mathbf{P}} X_{(j)}$$

INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
0	and the stars			

Example: for  $i \in \{1, 2\}$ , let

$$f_{i} = P_{1i} \times \mathcal{N}\left(M_{1i}, S_{1i}^{2}\right) + P_{2i} \times \mathcal{N}\left(M_{2i}, S_{2i}^{2}\right)$$

with  $M_{1i} \sim \mathcal{U}[-1.8, -1.2]$ ,  $M_{2i} \sim \mathcal{U}[1.2, 1.8]$ , n = 1000 and

Simulation 1:

- $P_{11} = P_{12} = 0.6$
- $P_{21} = P_{22} = 0.4$
- $S_{1i} \sim \mathcal{U}[0.2, 0.8]$
- $S_{2i} \sim \mathcal{U}[0.7, 1.3]$

Simulation 2:

- $P_{1i} \sim \mathcal{U}[0.45, 0.75]$
- $P_{2i} = 1 P_{1i}$
- $S_{11} = S_{12} = 0.5$
- $S_{21} = S_{22} = 1$

-

< ロ > < 同 > < 回 > < 回 > < 回 > <

Elie Maza
Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 000000
Quantile norm	alization			



Manifolds, a new framework for warped curves analysis

INP-ENSAT – INRA (laboratoire GBF)

<ロ> <同> <同> <同> <同> <同>

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 000000000	0000000 000000000 00000000 0000000
Quantile no	ormalization			



Manifolds, a new framework for warped curves analysis

INP-ENSAT – INRA (laboratoire GBF)

ヘロン 人間 と 人 ヨン 人 ヨン

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 000000000	0000000 000000000 00000000 0000000
0	and the second			

#### Quantile normalization

Elie Maza



## INP-ENSAT – INRA (laboratoire GBF)

▲ロト ▲圖ト ▲国ト ▲国ト

Plan	Microarray data analysis	Issue	<b>Some solutions</b> ○○○○○ ○○○○○○○○○●	New framework 0000000 00000000 00000000 0000000 00000
Quantile no	rmalization			
Limite	۹۲			

• Simulation 1:

$$F_i \approx F \circ H_i^{-1}$$

• Simulation 2:

Elie Maza

 $f_i \approx f \circ H_i^{-1}$ 

INP-ENSAT - INRA (laboratoire GBF)

メロト メポト メヨト メヨト

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 000 00000000	New framework 0000000 00000000 00000000 00000000 0000
Plan				
<ol> <li>Micro</li> <li>Issue</li> <li>Some         <ul> <li>Some</li> <li>Sh</li> <li>Ph</li> <li>Qu</li> </ul> </li> </ol>	oarray data analysis e solutions ift model nase model uantile normalization			
<ul> <li>4 New</li> <li>Ide</li> <li>Ap</li> <li>Sh</li> <li>Ap</li> </ul>	framework ea pproximation of $\delta$ ift model oplication		1011400143	

INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New framework ●000000 ○○○○○○○○ ○○○○○○○ ○○○○○○
Idea				
Plar	า			
<b>1</b> N	licroarray data analysis			
2 ls				
3 S	ome solutions Shift model Phase model Quantile normalization			
4 N	lew framework			
•	ldea			
٥	Approximation of $\delta$			
•	Shift model			
۲	Application			
				<ul> <li>▲ 王</li> <li>• ● ○ </li> <li>• ● ○ </li> </ul>
Elie Maz	za		INP-ENSAT – I	NRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Idea				



INP-ENSAT – INRA (laboratoire GBF)

**A** ► <

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Idea				



## INP-ENSAT – INRA (laboratoire GBF)

A B > 4
 A

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Idea				



## INP-ENSAT – INRA (laboratoire GBF)

Image: A math a math

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New framework 0000●00 00000000 00000000 0000000
ldea				

# Euclidean mean $\mu \in \mathbb{R}^n$ :

 $\hat{\mu} = \arg\min_{\mu \in \mathbb{R}^n} \sum_{i=1}^m \left[ \mathrm{d} \left( X^i, \mu \right) \right]^2$ 

INP-ENSAT - INRA (laboratoire GBF)

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 00000000 00000000 0000000
Idea				

# Euclidean mean $\mu \in \mathbb{R}^n$ :

$$\hat{\mu} = \arg\min_{\mu \in \mathbb{R}^n} \sum_{i=1}^m \left[ \mathrm{d} \left( X^i, \mu \right) \right]^2$$

Intrinsic mean  $\eta \in \mathcal{M}$  :

$$\hat{\eta} = \arg\min_{\eta \in \mathcal{M}} \sum_{i=1}^{m} \left[ \frac{\delta}{\delta} \left( X^{i}, \eta \right) \right]^{2}$$

INP-ENSAT - INRA (laboratoire GBF)

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 0000 00000000	New framework ○○○○●○○ ○○○○○○○○ ○○○○○○○ ○○○○○○
Idea				

Euclidean mean  $\mu \in \mathbb{R}^n$  :

$$\hat{\mu} = \arg\min_{\mu \in \mathbb{R}^n} \sum_{i=1}^m \left[ \mathrm{d} \left( X^i, \mu \right) \right]^2$$

Intrinsic mean  $\eta \in \mathcal{M}$  :

$$\hat{\eta} = \arg\min_{\eta \in \mathcal{M}} \sum_{i=1}^{m} \left[ \delta\left(X^{i}, \eta\right) \right]^{2}$$

Intrinsic mean  $\eta \in \mathcal{M}$  :

$$\hat{\eta} = \arg\min_{\eta \in \{X^1, \dots, X^m\}} \sum_{i=1}^m \left[ \hat{\delta} \left( X^i, \eta \right) \right]^2$$

INP-ENSAT - INRA (laboratoire GBF)

▲ @ ▶ ▲ ∃ ▶

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 0000000 000000
Idea				



INP-ENSAT – INRA (laboratoire GBF)

• • • • • • • • • • •

#### Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Idea				



INP-ENSAT – INRA (laboratoire GBF)

Image: A math a math

Manifolds, a new framework for warped curves analysis

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New framework ○○○○○○○ ●○○○○○○○○○ ○○○○○○○ ○○○○○○
Approximat	ion of $\delta$			
Plan				
1 Mic	roarray data analysis			
<ol> <li>Issu</li> </ol>				
3 Son • S • F • G	ne solutions hift model Phase model Quantile normalization			
<ul> <li>A</li> <li>A</li> <li>A</li> <li>A</li> <li>A</li> </ul>	v framework dea Approximation of $\delta$ whift model Application			
Elia Mara				► < E ► E ∽ Q @



-			
-			
		•••	

Issue

Some solutions

#### New framework

0000000 000000000 00000000 0000000

#### Approximation of $\delta$



# Euclidean complete graph

INP-ENSAT - INRA (laboratoire GBF)



Manifolds, a new framework for warped curves analysis

INP-ENSAT – INRA (laboratoire GBF)

	00000 000 00000000	0000000 000000000 00000000 0000000
Approximation of $\delta$		
	$\cup_{i=1}^{n}\mathcal{B}(\lambda)$	$(x_i, r_i)$
	◆□▶ ◆圖▶ ◆≧▶ ◆◎	। स्रिक्ट विक्रि
Elie Maza	INP-ENSAT – INRA (	(laboratoire GBF)

Some solutions

New framework

#### Elie Maza

Plan

Manifolds, a new framework for warped curves analysis

Microarray data analysis

1 ian		15500	00000 000 000000000	000000 00000000 00000000 0000000
Approximation of	δ			
			$\overline{X_iX_j} \in \cup_{k=1}^n$ or $\overline{X_iX_j} \notin \cup_{k=1}^n$	${}_{1}\mathcal{B}(X_{k},r_{k})$ ${}_{1}\mathcal{B}(X_{k},r_{k})$
Elie Maza			INP-ENSAT – INRA	(laboratoire GBF)

NL.

framowork

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 000 000000000	New framework ○○○○○○ ○○○○○○●○○ ○○○○○○○○ ○○○○○○
Approximatio	on of $\delta$			
			n = 30 obs	erved points
Elie Maza			INP-ENSAT – IN	IRA (laboratoire GBF)

-		
L		

Issue

Some solutions

New framework

0000000 000000000 00000000 0000000

#### Approximation of $\delta$



# n = 100 observed points

INP-ENSAT - INRA (laboratoire GBF)

• • • • • • • • • • • •

n	I				
г	l	d	I	I	

Issue

Some solutions

#### New framework

0000000 00000000 00000000 00000000

#### Approximation of $\delta$



# n = 300 observed points

INP-ENSAT - INRA (laboratoire GBF)

• • • • • • • • • • •

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New framework ○○○○○○○ ○○○○○○○○ ●○○○○○○○ ○○○○○○
Shift mode	el			
Plan				
1 Mi	croarray data analysis			
2 Iss				
3 Sol • S • I • (	me solutions Shift model Phase model Quantile normalization			
<ul> <li>4 Ne</li> <li>1</li> <li>7</li> </ul>	w framework dea Approximation of $\delta$ Shift model			
• /	Application			
			< <b>□ &gt; &lt;□</b> > < ≣	<ul> <li>▲ Ξ</li> <li>◆ Ξ</li></ul>
Elie Maza			INP-ENSAT – I	NRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 00000 000 000000000	New framework ○○○○○○○ ○●○○○○○○ ○●○○○○○○
Shift model				

## Model:

$$X_i^j = f(t_j - A_i), \ i \in \{1, \dots, n\}, \ j \in \{1, \dots, m\}$$

with

- $f:\mathbb{R}\to\mathbb{R}$  an unknown function
- A an unknown real valued variable
- $A_i \sim A$  iid

• 
$$t_j \in \mathbb{R}$$

Image: A math the second se



# We define the structural median of the stochastic function $f(\cdot - A)$ by

$$f_{\mathrm{SM}} = f\left(\cdot - \mathrm{med}(A)\right)$$

A natural estimator of  $f_{\rm SM}$  is

$$\widehat{f}_{\mathrm{SM}} = \left( f\left(t_1 - \widehat{\mathrm{med}}(A)\right), f\left(t_2 - \widehat{\mathrm{med}}(A)\right), \dots, f\left(t_m - \widehat{\mathrm{med}}(A)\right) \right)$$

INP-ENSAT - INRA (laboratoire GBF)

<ロト </p>

Elie Maza

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 0000000
Shift model				

## Let

Elie Maza

$$\begin{array}{rcl} X:\mathbb{R} & \rightarrow & \mathbb{R}^m \\ a & \mapsto & X(a) = \left(f\left(t_1-a\right), f\left(t_2-a\right), \ldots, f\left(t_m-a\right)\right) \end{array}$$

## Lemma: The set

$$\mathcal{C} = \{X(a) \in \mathbb{R}^m, \ a \in \mathbb{R}\}$$

is a 1-dimensional embedded manifold, with distance

$$\delta(X_1, X_2) = \left| \int_{a_1}^{a_2} \left\| X'(a) \right\| \mathrm{d}a \right|$$

where  $X_1 = X(a_1)$  and  $X_2 = X(a_2)$ .

INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 0000000 000000
Shift model				

Theorem: we have 
$$\widehat{\mu}_{I}^{1} = \widehat{f}_{SM}$$
.

INP-ENSAT – INRA (laboratoire GBF)

3

・ロト ・ 日 ト ・ ヨ ト ・ ヨ ト

Manifolds, a new framework for warped curves analysis

Plan	_		
PIAR			
1 10111			

Issue

Some solutions

#### New framework

0000000 000000000 00000000 00000000

#### Shift model

Elie Maza



## Example:

• 
$$f(t) = \exp\left(-t^2\right)$$

• 
$$A \sim U(] - 1, 1[)$$

• 
$$t_1 = 0.5, t_2 = 1$$

INP-ENSAT - INRA (laboratoire GBF)

・ロト ・回ト ・ヨト ・ヨ

Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 00000000
Shift model				



INP-ENSAT – INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	<b>Some solutions</b> 0000 000000000	New framework ○○○○○○○ ○○○○○○○ ○○○○○○● ○○○○○○
Shift model				



INP-ENSAT - INRA (laboratoire GBF)

Plan	Microarray data analysis	Issue	Some solutions 00000 000 00000000	New framework ○○○○○○○ ○○○○○○○○ ○○○○○○○
Applicati	ion			
Plar	ı			
<b>1</b> M	licroarray data analysis			
<b>2</b> ls				
3 So • •	ome solutions Shift model Phase model Quantile normalization			
4 N	ew framework			
0	Idea			
۲	Approximation of $\delta$			
۲	Shift model			
۹	Application			
			<ul><li>&lt; □ &gt; &lt; □ &gt; &lt; Ξ</li></ul>	<ul><li>&lt; ≣ &lt; ≤ &lt; &lt; </li></ul>
Elie Maz	za		INP-ENSAT – I	NRA (laboratoire GBF)



Manifolds, a new framework for warped curves analysis







Manifolds, a new framework for warped curves analysis
Plan	Microarray data analysis	Issue	Some solutions	New framework
			00000 000 00000000	0000000 000000000 00000000 <b>00000</b> ●
Application				

## Thank you for your attention!

Elie Maza

INP-ENSAT - INRA (laboratoire GBF)

Manifolds, a new framework for warped curves analysis