Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics

Marie Doumic (BANG project-team – INRIA & Paris VI)

Rocquencourt – May 28th, 2008

with B. PERTHAME (L. J-L. Lions-Paris)
J. ZUBELLI (IMPA-Rio de Janeiro)
and Pedro MAIA (UFRJ - master student)

Outline

Structured Population Models

Motivation to structured population models

The model under consideration

The Inverse Problem

Regularization by quasi-reversibility method

Regularization by filtering approach

Numerical Solution

Choice of a convenient scheme

Some results and application to experimental data

Conclusion and perspectives

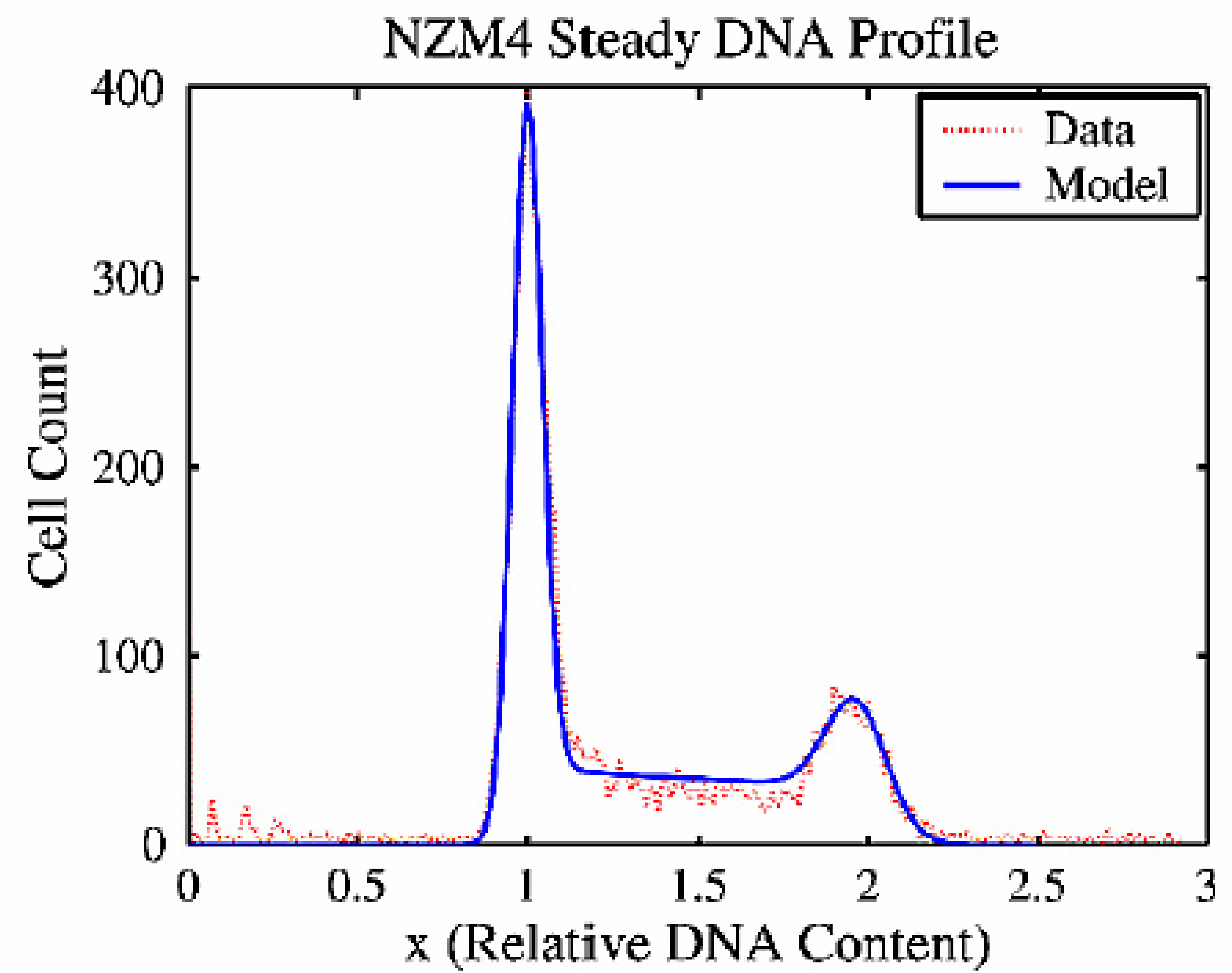
Structured Populations

Recent Ref: B. Perthame, *Transport Equations in Biology,* Birkhäuser (2006)

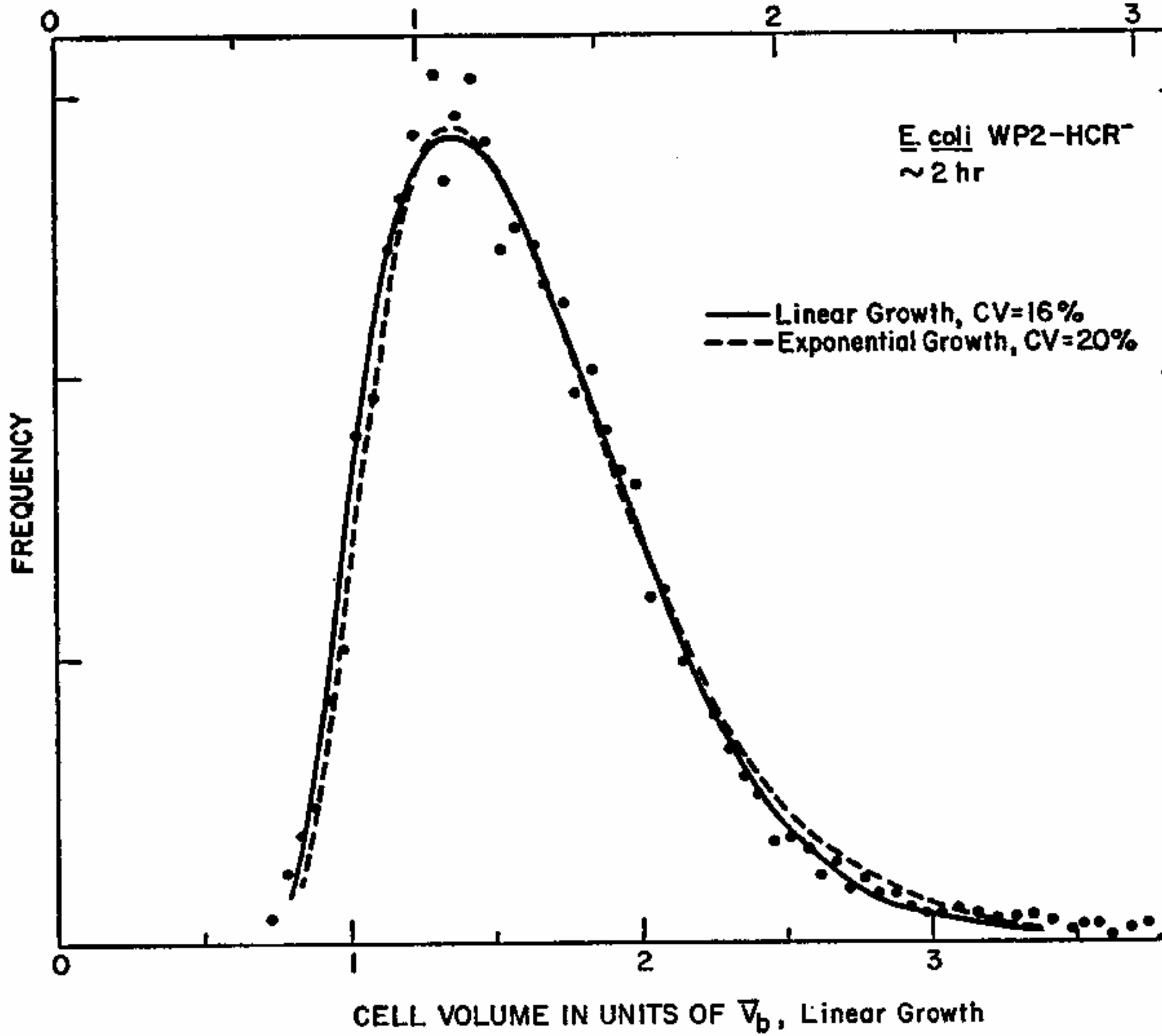
Population density $n = n(t, x)$

Examples of *x*:

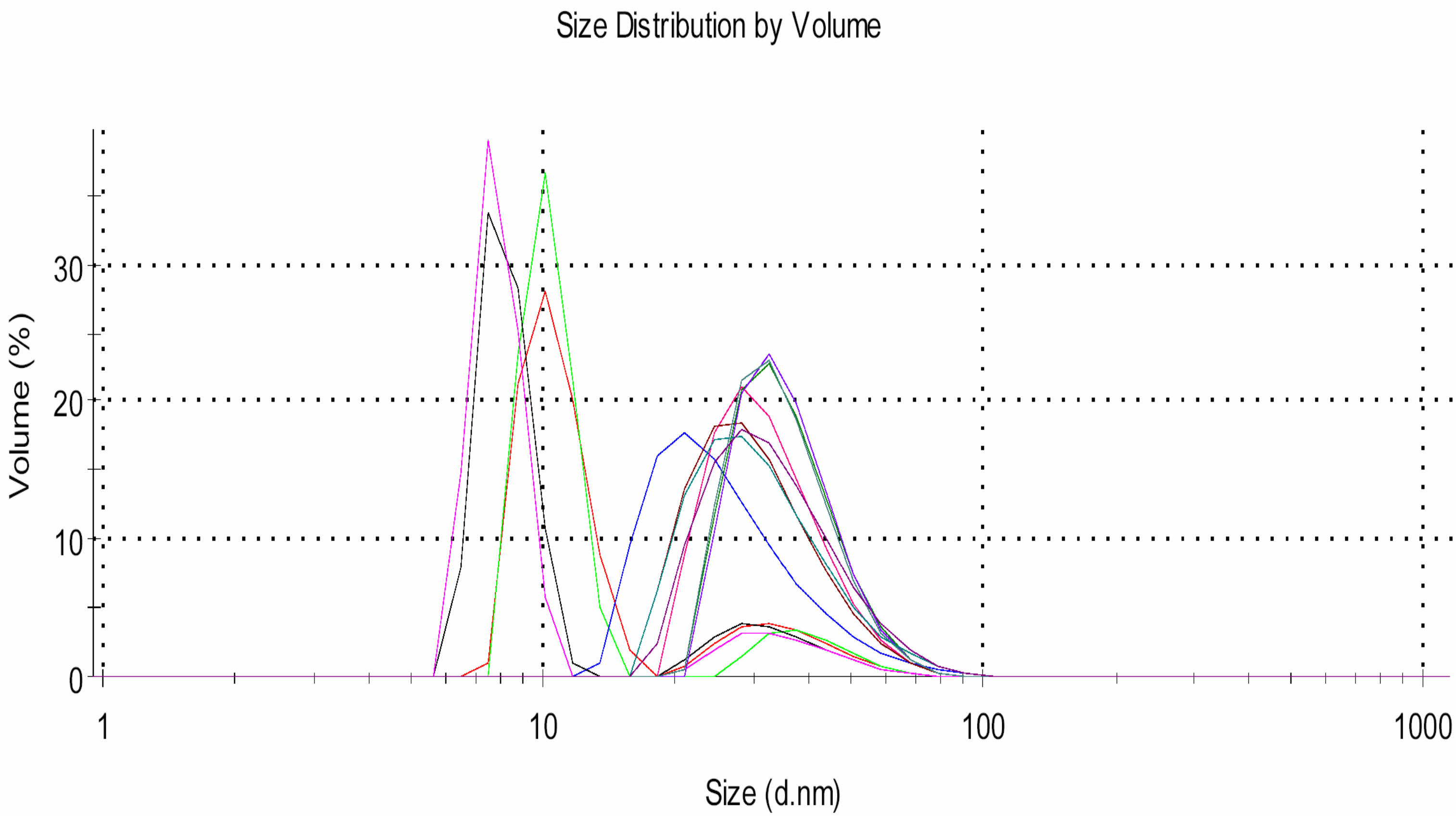
- Unicellular organisms: the mass of the cell
- DNA content of the cell
- Cell age (age-structured populations)
- Protein content: cyclin, cyclin-dependent kinases, complexes…
- If n is a density of polymers: size of the polymer



From B. Basse *et al, Modeling the flow of cytometric data obtained from unperturbed human tumor cell lines: parameter fitting and comparison*. of Math. Bio., 2005



Cell volumes distribution for E. Coli THU in a glucose minimal medium at a doubling time of 2 hrs. H.E. Kubitschek, Biophysical J. 9:792-809 (1969)



Size distribution kinetic of PrP polymerization in physico-chemical conditions
leading to the formation of amyloid fibrils monitored by MWSLS technic (taken from ANR TOPPAZ, INRA/BPCP data).

Marie DOUMIC (INRIA, BANG)
Nice, September 14th, 2009

Our Structured Population Model: Model of Cell Division under Mitosis

Density of cells: $n(t, x)$

Size of the cell: x

Birth rate: $B(x)$

1 cell of size $2x$ gives birth to 2 cells of size x

The growth of the cell size by nutrient uptake is given by a rate *g(x)*:

here for the sake of simplicity, $g(x) \equiv 1$.

Our Structured Population Model: Model of Cell Division under Mitosis

Model obtained by a mass conservation law:

$$
\begin{cases}
\frac{\partial}{\partial t}n(t,x) + \frac{\partial}{\partial x}n(t,x) + B(x)n(t,x) = 4B(2x)n(t,2x), & x \geqslant 0,\ t \geqslant 0, \\
n(t, x=0) = 0,\ t > 0, \\
n(0,x) = n^0(x) \geq 0.
\end{cases}
$$

Density of the cells

Growth by nutrient

Division of cells of size *x*

Division of cells of size *2x*

Applications of this model (or adaptations of this model): cell division cycle, prion replication, fragmentation equation…

(Some) Related Works

- J.A.J Metz and O. Diekmann,

  Physiologically Structured Models (1986)

- Engl, Rundell, Scherzer,

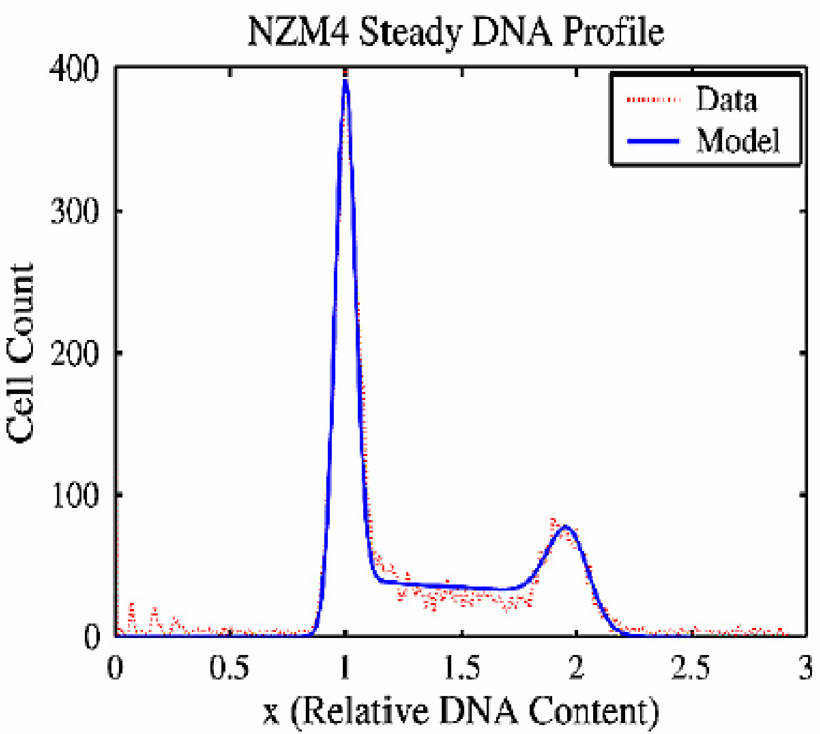
  Regularization Scheme for an Inverse Problem in Age-Structured Populations (1994)

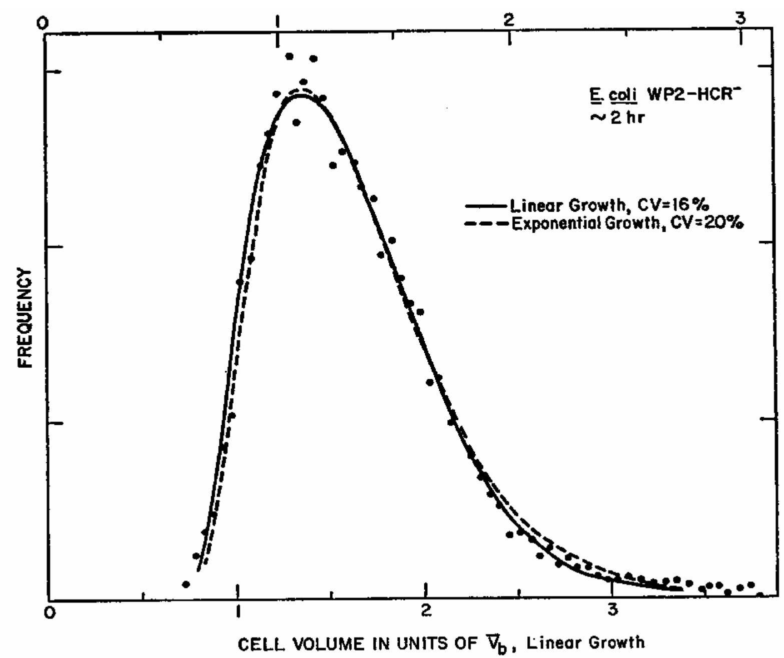
- Gyllenberg, Osipov, Päivärinta,

  The Inverse Problem of Linear Age-Structured Population Dynamics (2002)

The Question: find the birth rate *B*

What is really observed ? Recall the figures:





⇒ We do not observe *B* ; not even *n(t,x)*:

but a DOUBLING TIME and a STEADY PROFILE *N(x)*

Our approach:
Use the *stable size distribution*

This uses recent results on Generalized Relative Entropy: refer for instance to

- B. Perthame, L. Ryzhik, *Exponential Decay for the fragmentation or cell-division Equation* J. Diff. Equ. (2005)

- P. Michel, S. Mischler, B. Perthame, General Relative Entropy for Structured Population Models and Scattering, C.R. Math. Acad. Sci. Paris (2004)

Dynamics of this equation

If you look at solutions n(t,x) under the form *n(t,x)=e^λt N(x):*

Theorem (above ref.): There is a unique solution $N(x)$ for a unique $\lambda_0 > 0$ of:

$$\begin{cases} \frac{\partial}{\partial x} N + (\lambda_0 + B(x))N = 4B(2x)N(2x), & x \geq 0, \\ N(x=0) = 0, \\ N(x) > 0 \text{ for } x > 0, \quad \int_0^\infty N(x)dx = 1. \end{cases}$$

And under fairly general conditions we have (in weighted L^p topologies) :

$$n(t,x)e^{-\lambda_0 t} \xrightarrow[t\to\infty]{} \rho N(x)$$

(Assumptions on *B(x)* quite general – exponential decay can also been proved)

Dynamics of the equation

2 Major & fundamental & useful relations:

1. Integration of the equation:

$$\lambda_0 = \int B(x)N(x)dx.$$

Interpretation: number of cells increases by division

2. Integration of the equation multiplied by *x*:

$$\lambda_0 \int xN(x)dx = 1.$$

Interpretation: biomass increases by nutrient uptake

1st question on our inverse problem: is it well-posed ?

Hadamard's definition of well-posedness:

1- For all admissible data, a solution exists

2- For all admissible data, the solution is unique

3- The solution depends continuously on the data.

Our problem becomes: Knowing $N \geq 0$, with $N(x=0)=0$, find B such that :

$$\frac{\partial}{\partial x} N + (\lambda_0 + B(x))N = 4B(2x)N(2x)$$

An ill-posed problem

Now N is the parameter, B the unknown.
The equation can be written as:

$$4B(2x)N(2x) - B(x)N(x) = L(x),$$

With $L(x) = \frac{\partial}{\partial x} N(x) + \lambda_0 N(x)$

If N is regular, e.g. if L is in L^2 : $H=BN$ is in L^2 (see prop. below)
what we really know is not N but a noisy data N_ε with

$\|N_\varepsilon - N\|_{L^2} \leq \varepsilon$ ⟹ L is **not** in L^2
B is **not** well-defined in L^2

How to regularize this problem: « quasi-reversibility » method

1st method (in Perthame-Zubelli, *Inverse Problems* (2006)):

Add a (small) derivative for B: we obtain the following well-posed problem:

$$4B_{\epsilon,\alpha}(2x)N_\epsilon(2x) - B_{\epsilon,\alpha}(x)N_\epsilon(x) = \frac{\partial}{\partial x}N_\epsilon(x) + \lambda_\epsilon N_\epsilon(x), \quad x > 0,$$

$$(B_{\epsilon,\alpha}N_\epsilon)(0) = 0.$$

Theorem (Perthame-Zubelli): we have the error estimate:

$$\|B_{\varepsilon,\alpha} - B\|_{L^2(N_\varepsilon^2 dx)} \leq C\alpha\|N\|_{H^2} + \frac{C + \|B\|_{L^\infty}}{\alpha}\|N_\varepsilon - N\|_{L^2}.$$

$\Longrightarrow$ $\alpha = O(\sqrt{\epsilon})$ is optimal

How to regularize this problem – Filtering approach

2nd method: regularize $\frac{\partial}{\partial x}N$ in order to make *L* be regular:

$$4B_{\epsilon,\alpha}(2x)N_\epsilon(2x) - B_{\epsilon,\alpha}(x)N_\epsilon(x) = \left(\frac{\partial}{\partial x}N_\epsilon + \lambda_{\epsilon,\alpha}N_\epsilon\right) * \rho_\alpha(x) \qquad x > 0,$$

$$(B_{\epsilon,\alpha}N_\epsilon)(0) = 0.$$

With $\rho_\alpha(x) = \frac{1}{\alpha}\rho(\frac{x}{\alpha})$, and $\rho \in \mathcal{C}_c^\infty(\mathbb{R})$

Theorem (D-Perthame-Zubelli, *Inverse Problems*, 2009): we have the error estimate:

$$||B_{\epsilon,\alpha} - B||_{L^2(N_\epsilon^2 dx)} \leqslant C(\alpha + \sqrt{|\lambda_\epsilon - \lambda_0|})||N||_{H^2(\mathbb{R}_+)} + \frac{C + ||B||_{L^\infty(\mathbb{R}_+)}}{\alpha}||N_\epsilon - N||_{L^2(\mathbb{R}_+)}.$$

$\Longrightarrow$ $\alpha = O(\sqrt{\epsilon})$ is optimal

Numerical Implementation

Generic form of the problem for any regularization:

$$4B(2x)N(2x) - B(x)N(x) = L(x),$$

With 1.Naive method:

$$L(x) = \frac{\partial}{\partial x} N_\epsilon(x) + \lambda_\epsilon N_\epsilon(x)$$

2. « Quasi-reversibility » method:

$$L(x) = \frac{\partial}{\partial x} N_\epsilon(x) + \lambda_{\epsilon,\alpha} N_\epsilon(x) - \alpha \frac{\partial}{\partial x}(B_{\epsilon,\alpha} N_\epsilon)(2x)$$

3. Filtering method:

$$L(x) = (\frac{\partial}{\partial x} N_\epsilon + \lambda_{\epsilon,\alpha} N_\epsilon) * \rho_\alpha(x)$$

Numerical Implementation - strategy

Several requirements:

A. Avoid instability

B. Conserve main properties of the continuous model: laws for the increase

- of biomass

$$\lambda_{\epsilon,\alpha} \int N_\epsilon dx = \int B_{\epsilon,\alpha} N_\epsilon dx,$$

- of number of cells, e.g. for the quasi-reversibility method:

$$-\alpha \int B_{\epsilon,\alpha} N_\epsilon dx = 4\lambda_{\epsilon,\alpha} \int x N_\epsilon dx - 4 \int N_\epsilon dx.$$

C. 1 question: begin from the left, deducing *B(2x)* from *B(x)* or from the right, deducing *B(x)* from *B(2x)* ?

$$4B(2x)N(2x) - B(x)N(x) = L(x),$$

Numerical Implementation: a result to choose the right scheme

Proposition. Let us consider the following equation:

$$4H(2x) - H(x) = L.$$

1) If $L = 0$, all solutions in $\mathcal{D}'(\mathbb{R}_+^*)$ are given by $\frac{f(Log(x))}{x^2}$ where $f \in \mathcal{D}'(\mathbb{R})$ is $Log(2)-$ periodic.

2) It admits *formally* the two following solutions:

$$H_1(x) = \sum_{n=1}^{+\infty} 2^{-2n} L(2^{-n}x), \qquad H_0(x) = -\sum_{n=0}^{+\infty} 2^{2n} L(2^n x).$$

3) If $L \in L^2(x^p dx)$ with $p \neq 3$, $\exists\,!$ solution $H \in L^2(x^p dx)$.

4) For $p < 3$, $H_1 \in L^2(x^p dx)$, and for $p > 3$, $H_0 \in L^2(x^p dx)$.

5) For $1 \leqslant q \leqslant \infty$, if $L \in L^q(\mathbb{R}_+)$ then $H_1 \in L^q(\mathbb{R}_+)$.

Numerical Implementation: choice of the scheme

$$H_1(x) = \sum_{n=1}^{+\infty} 2^{-2n} L(2^{-n}x), \qquad H_0(x) = -\sum_{n=0}^{+\infty} 2^{2n} L(2^n x).$$

H_0: deduce $H(x)$ from larger x ⟺ scheme departs from infinity

H_1: deduce $H(x)$ from smaller x ⟺ scheme departs from *0*.

H_1 « more regular » ⟹ choice: scheme departing from *0.*

Numerical Scheme – filtering approach

-> departs from zero (mimics H_1)

-> mass and number of cells balance laws preserved:

$$\lambda_{\epsilon,\alpha} = \frac{\int N_{\epsilon,\alpha}(x)dx}{\int xN_{\epsilon,\alpha}(x)dx}.$$

-> stability: $4H(2x)$ is approximated by $4\,H_{2i}$

$$\forall\, 0 \leqslant i \leqslant I, \quad 4H_i^f = H_{\frac{i}{2}}^f + L_{\frac{i}{2}}^f,$$

$$\forall\, 0 \leqslant i \leqslant I, \quad G_{\frac{i}{2}} = \begin{cases} G_{\frac{i}{2}} & \text{if } i \text{ is even,} \\ \frac{1}{2}\left(G_{\frac{i-1}{2}} + G_{\frac{i+1}{2}}\right) & \text{if } i \text{ is odd.} \end{cases}$$

Numerical Scheme – Quasi-Reversibility

-> departs from zero (mimics H_1)

-> mass and number of cells balance laws preserved:

$$\lambda_{\epsilon,\alpha} = \frac{\int N_\epsilon dx}{\int x N_\epsilon dx + \frac{\alpha}{4}\int N_\epsilon dx}$$

-> stability: $4H(2x)$ is approximated by $4\,H_{2i}$

$$\frac{\alpha}{\Delta x}(H_i^Q - H_{i-1}^Q) + 4H_i^Q = H_{\frac{i}{2}}^Q + L_{\frac{i}{2}}^Q,$$

$$L_i^Q = \lambda_\epsilon N_i + \frac{N_{i+1} - N_i}{\Delta x}.$$

Numerical Scheme: steps

1. solve the direct problem for a given *B(x)*

Method: use of the exponential convergence of *n(t,x)* to *N(x):*
Finite volume scheme to solve the time-dependent problem:

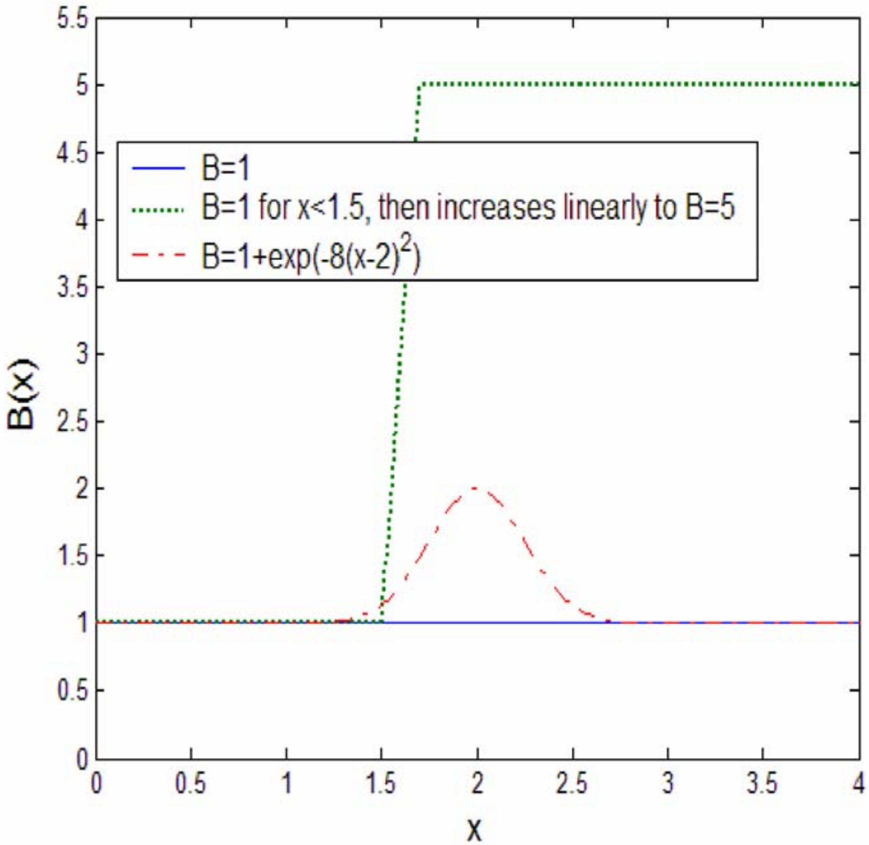
$$\frac{n_i^{k+1}-n_i^k}{\Delta t}+\frac{n_i^k-n_{i-1}^k}{\Delta x}+B_i n_i^{k+1}=B_{2i-1}n_{2i-1}^k+2B_{2i}n_{2i}^k+B_{2i+1}n_{2i+1}^k$$

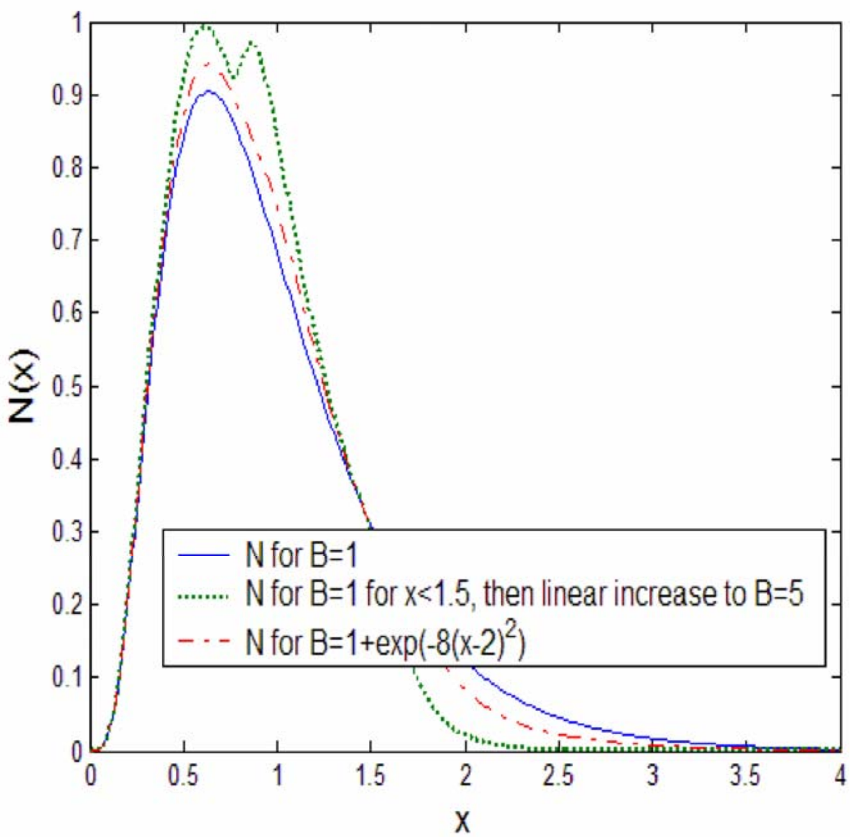
Then renormalization at each time-step

2. Add a noise to *N(x)* to get a noisy data $N_\varepsilon(x)$

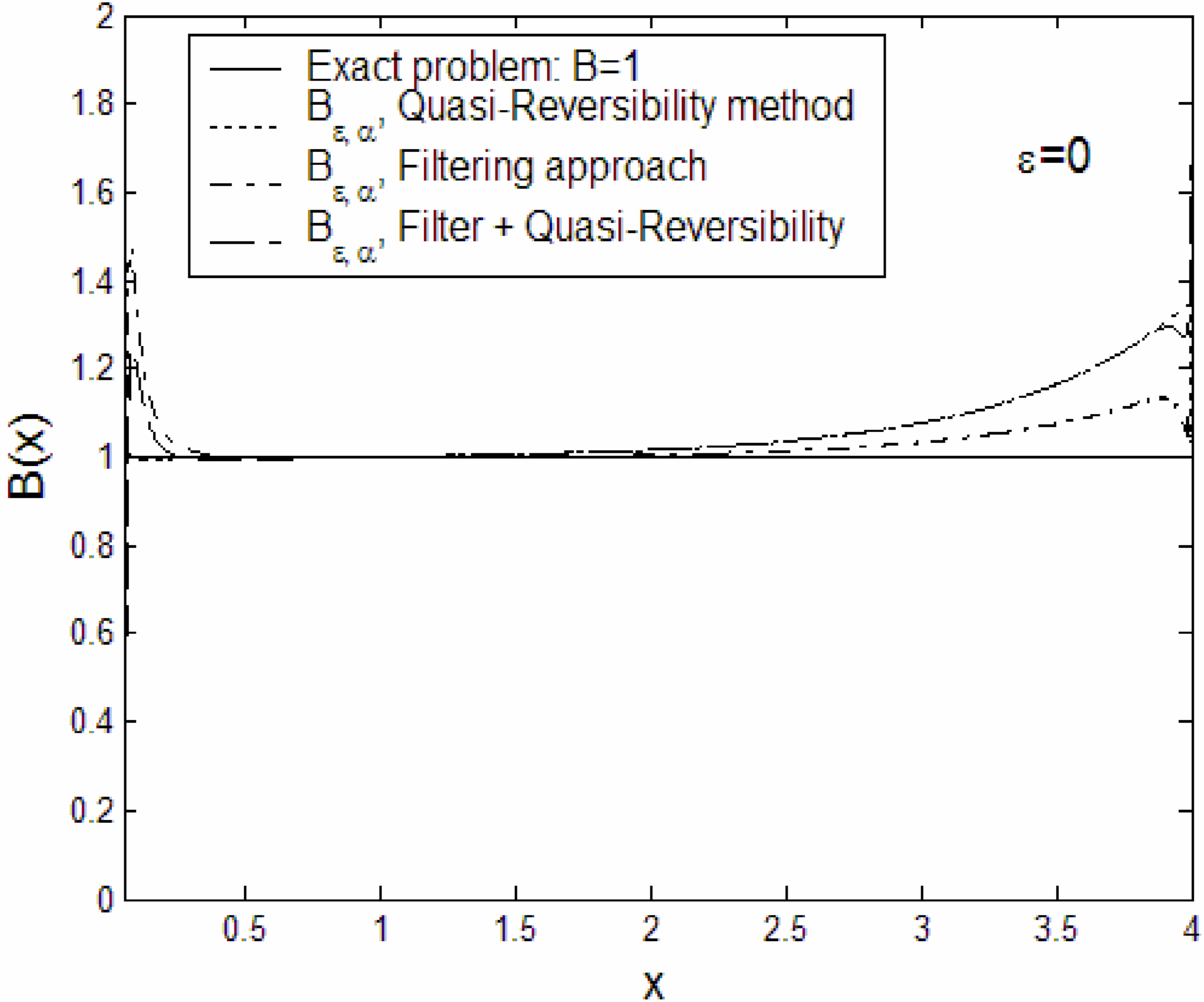
3. Run the numerical scheme for the inverse problem to get a birth rate $B_{\varepsilon,\alpha}(x)\, N_\varepsilon(x)$ and compare it with the initial data *B(x)* – *look for the best α for a given error ε*

First step: direct problem

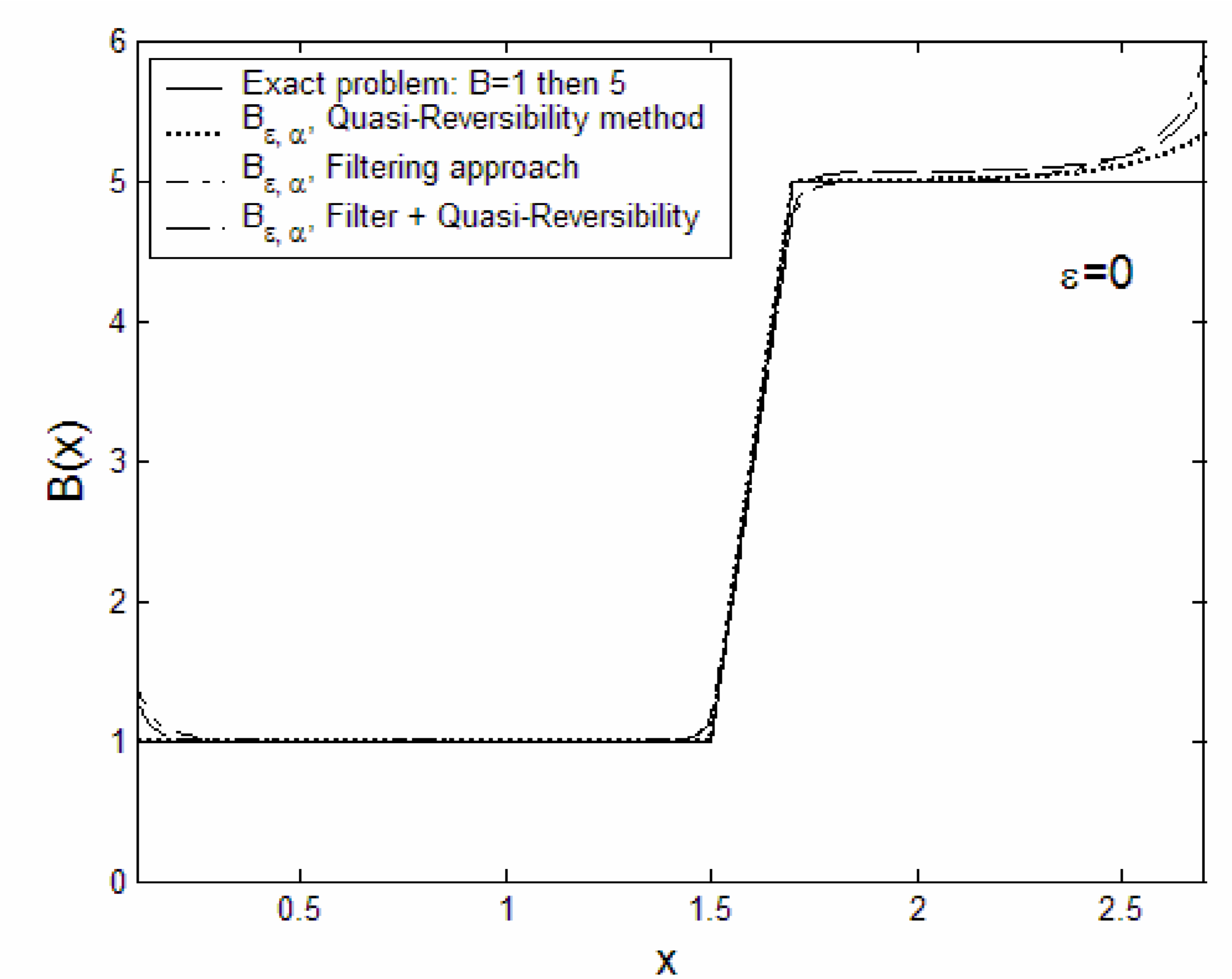




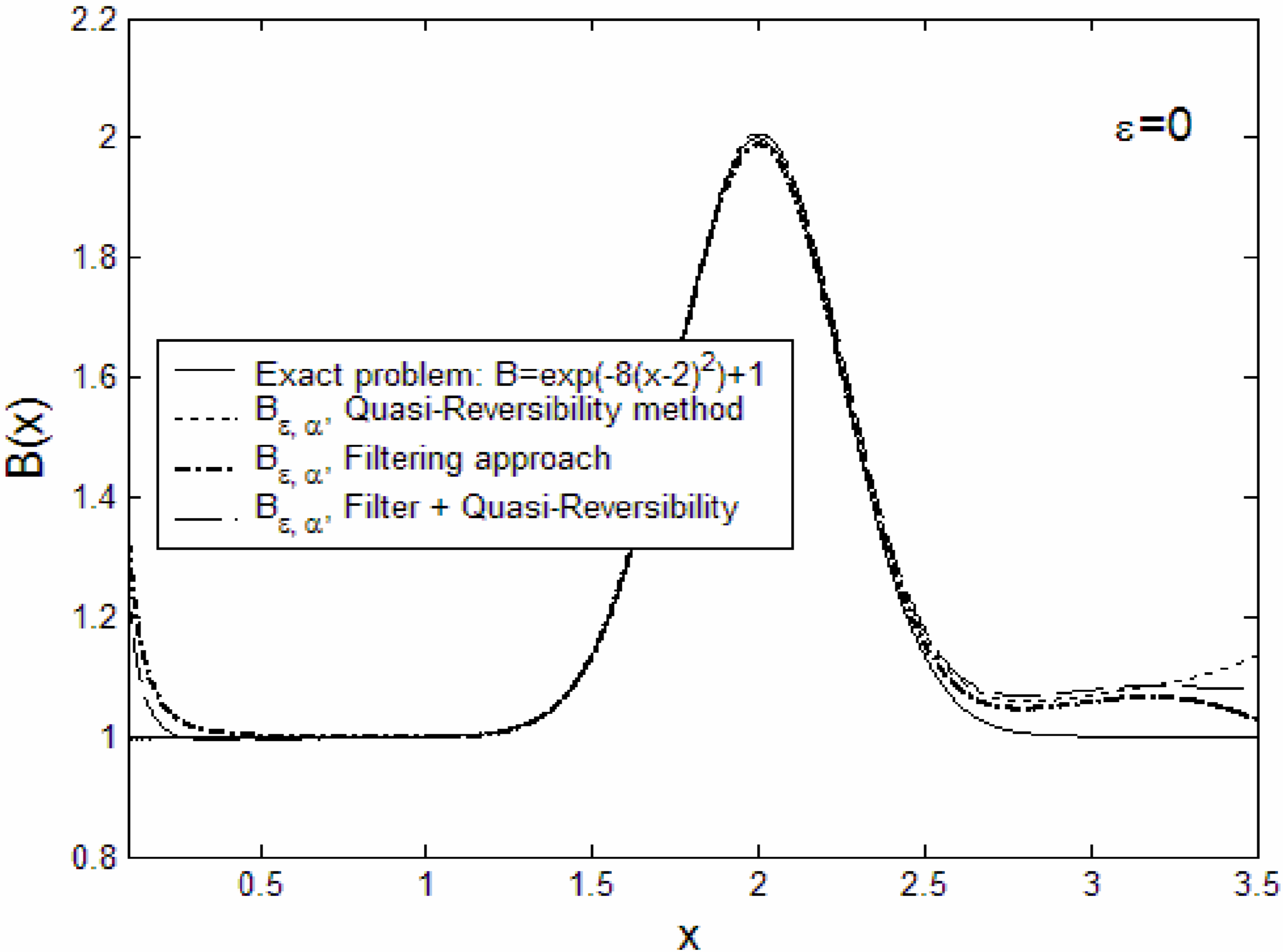
Results with ε=0: no noise for the entry data



Results with ε=0: no noise for the entry data

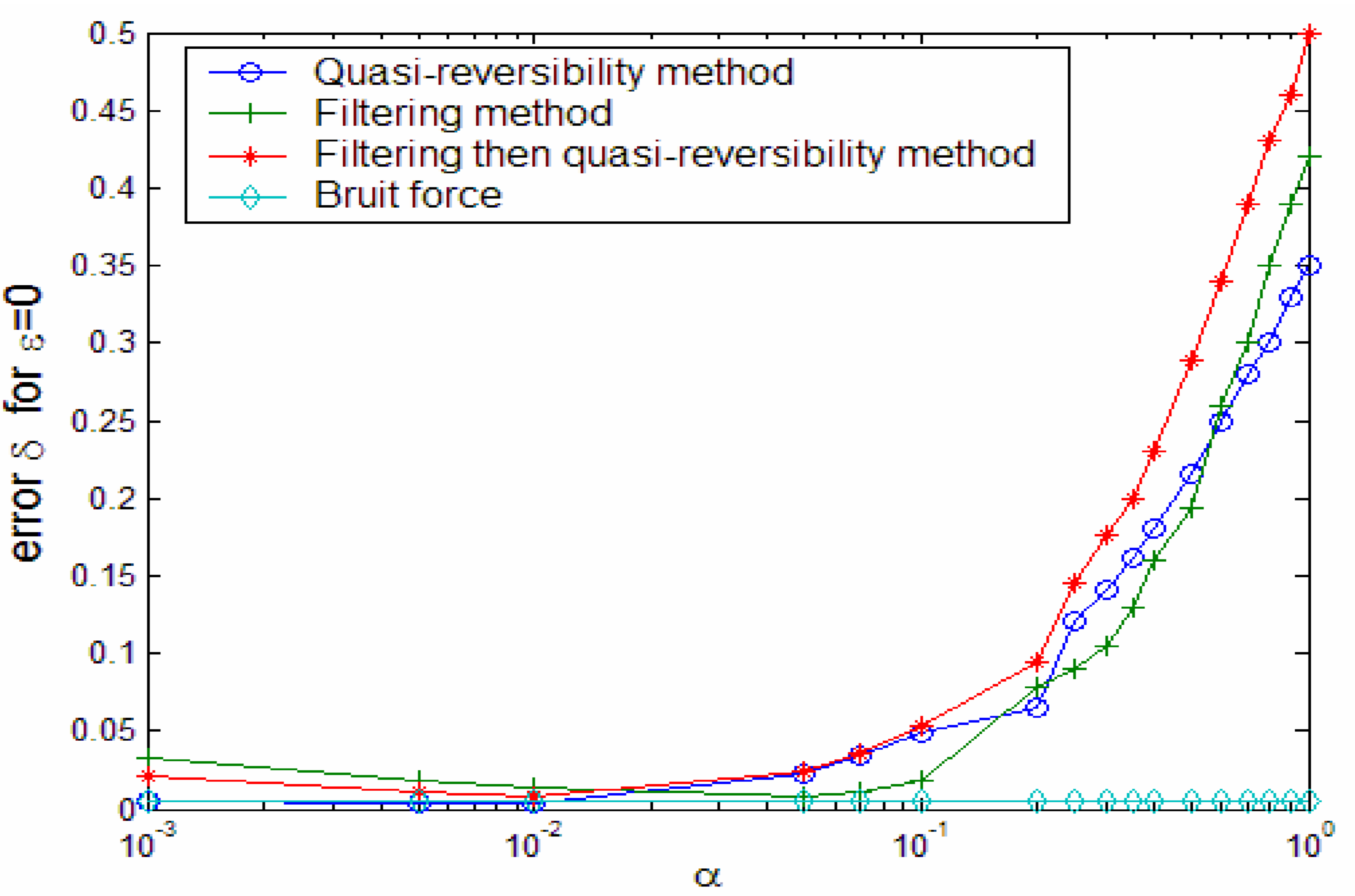


Results with ε=0: no noise for the entry data

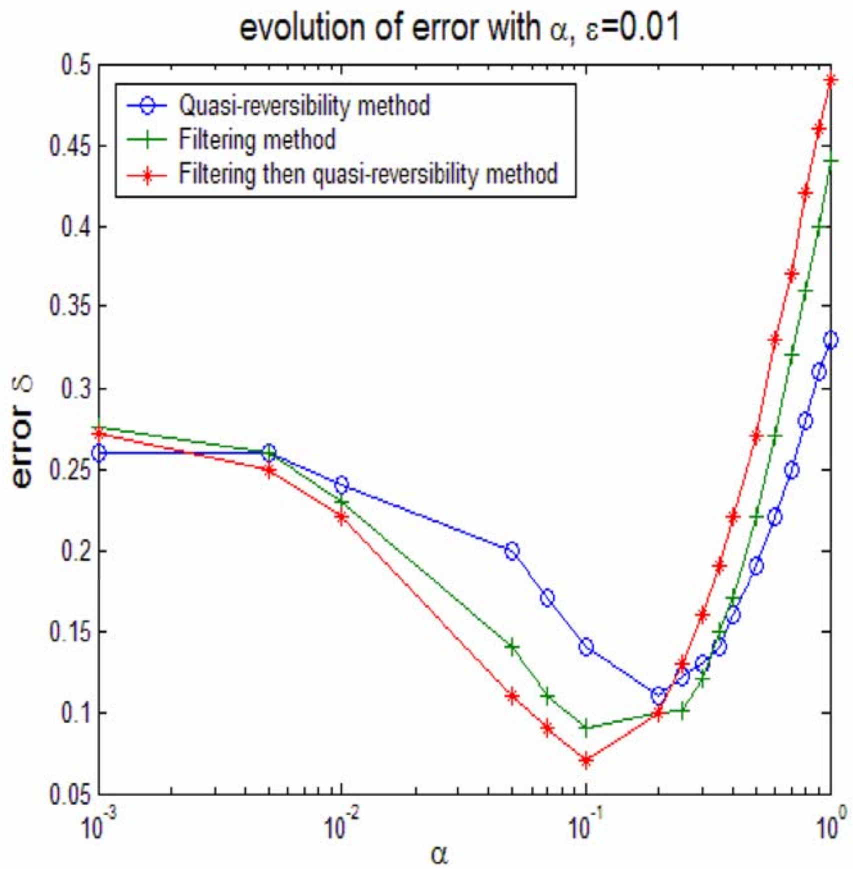


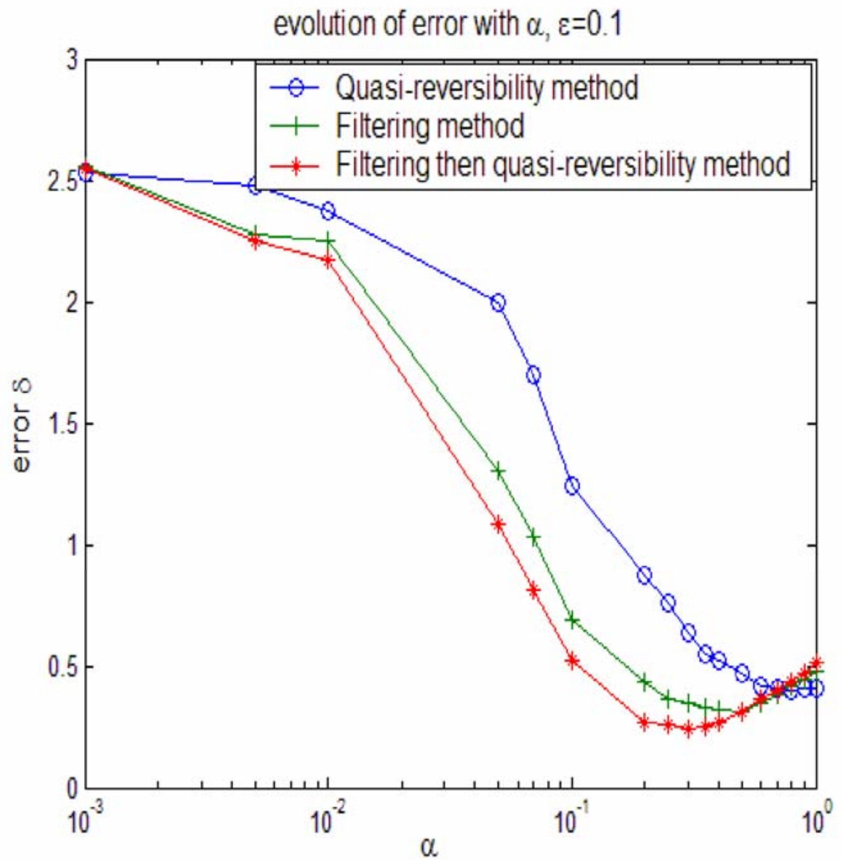
Results with ε=0: no noise for the entry data

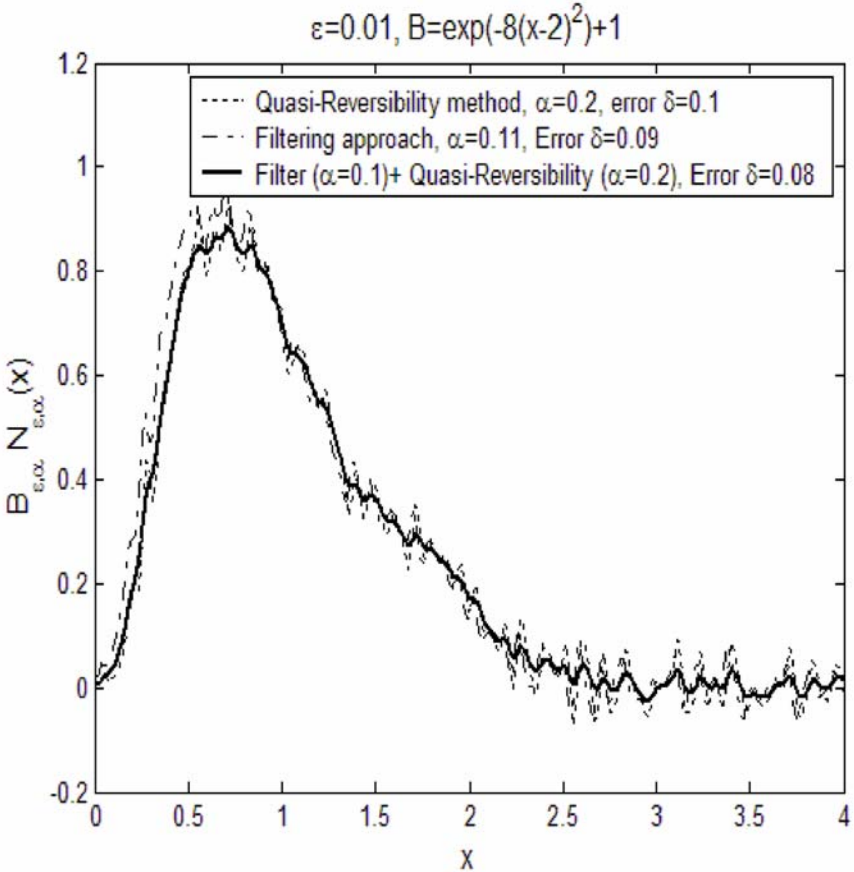
Measures of error for the different methods



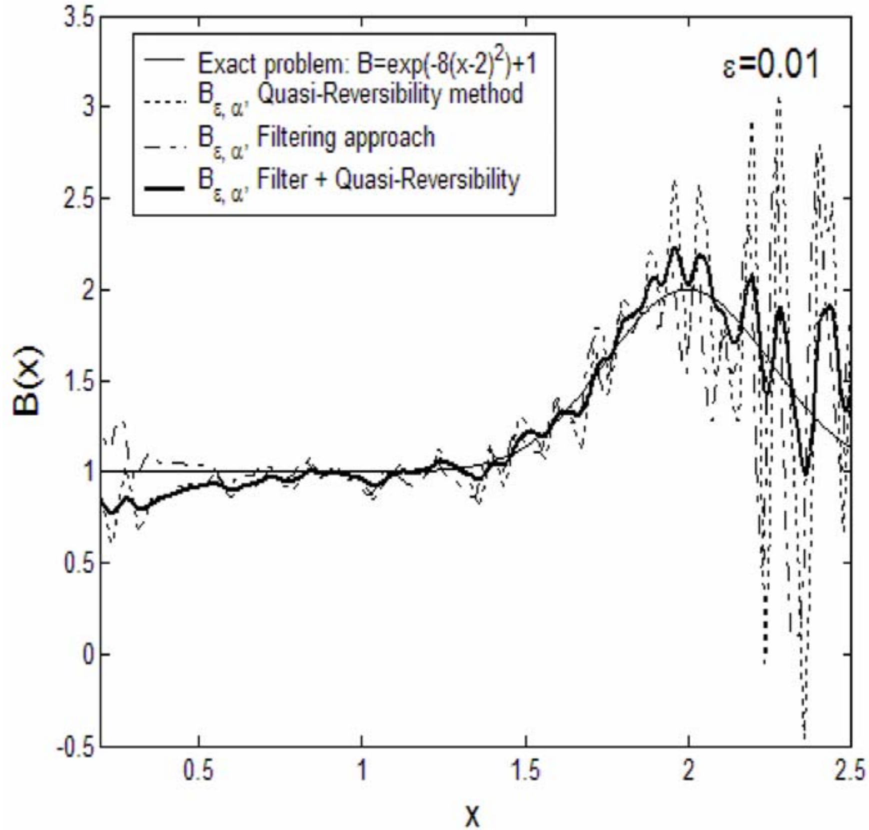
Results with ε=0.01 and 0.1: measure of error



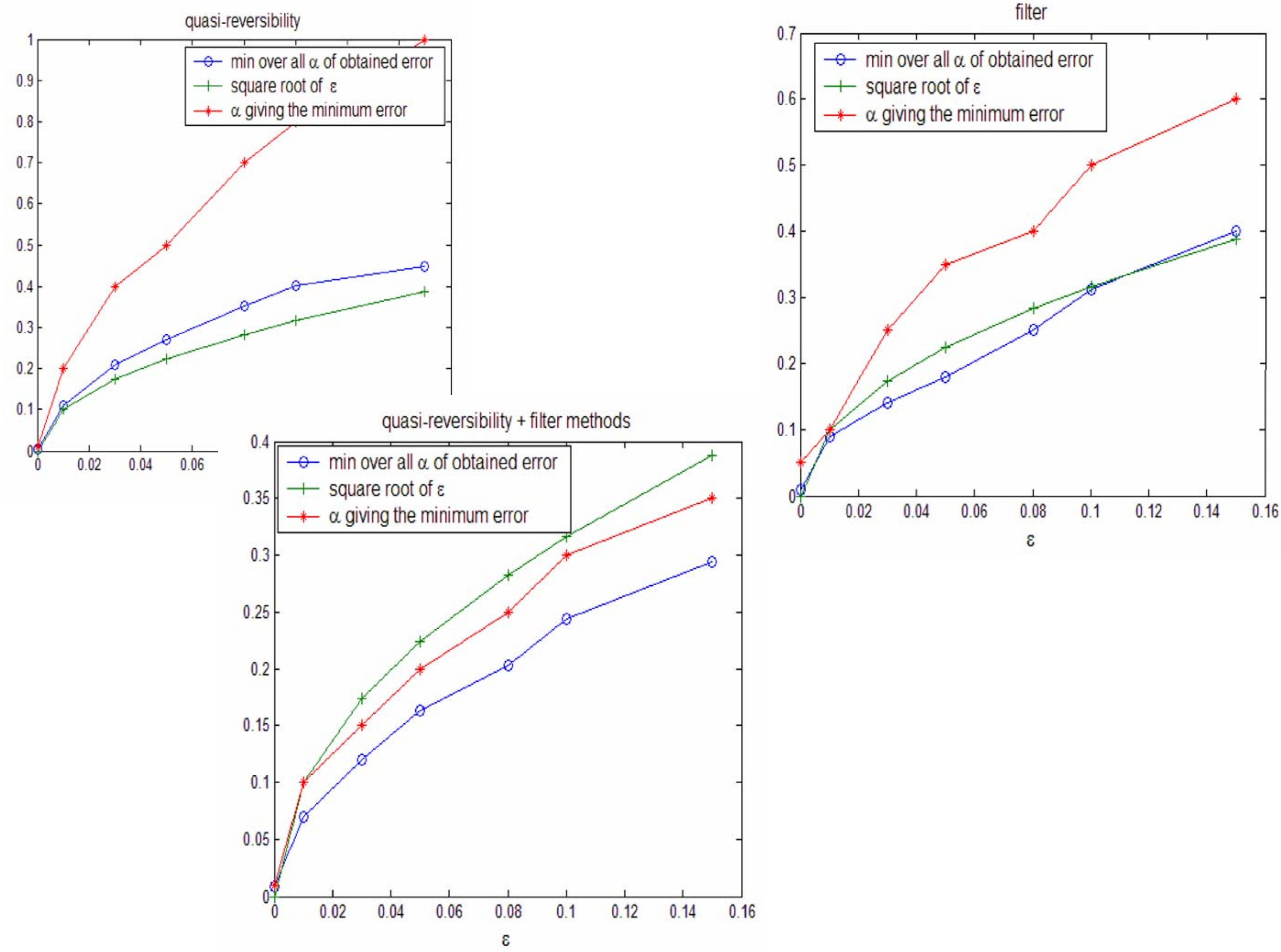




ε=0.01, B=exp(-8(x-2)^2)+1
Quasi-Reversibility method, α=0.2, error δ=0.1
Filtering approach, α=0.11, Error δ=0.09
Filter (α=0.1)+ Quasi-Reversibility (α=0.2), Error δ=0.08
B_{ε,α} N_{ε,α}(x)
x



Exact problem: B=exp(-8(x-2)^2)+1
B_{ε, α}, Quasi-Reversibility method
B_{ε, α}, Filtering approach
B_{ε, α}, Filter + Quasi-Reversibility
ε=0.01
B(x)
x



quasi-reversibility
min over all α of obtained error
square root of ε
α giving the minimum error
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
0 0.02 0.04 0.06
quasi-reversibility + filter methods
min over all α of obtained error
square root of ε
α giving the minimum error
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4
0 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16
ε
filter
min over all α of obtained error
square root of ε
α giving the minimum error
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
0 0.02 0.04 0.06 0.08 0.1 0.12 0.14 0.16
ε

Application to experimental data

(D-Maia-Zubelli, Proc. ECMTB, submitted, 2008)

Main difficulty: find data to which the equation can be applied !

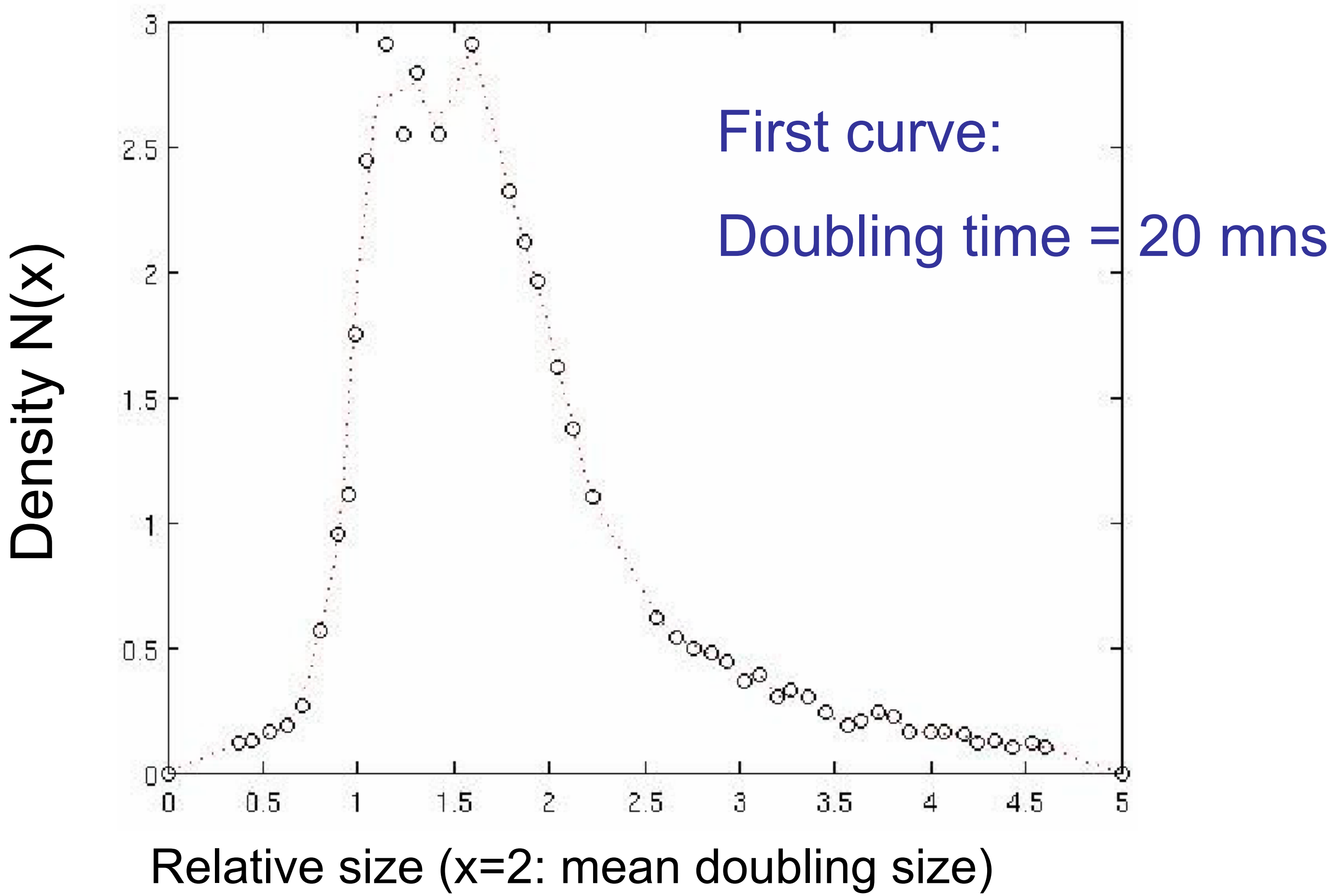
- No death (in vitro experiment) or constant death rate
- Symmetrical division
- Known growth speed (assumption is needed)

⇨ Bacteria: E. Coli

Reference: H.E. Kubitschek, *Growth during the bacterial cell cycle: analysis of cell size distribution,* Biophys. J., (1969)

Application to Kubitschek's data

4 kinds of growth environment for E. Coli:



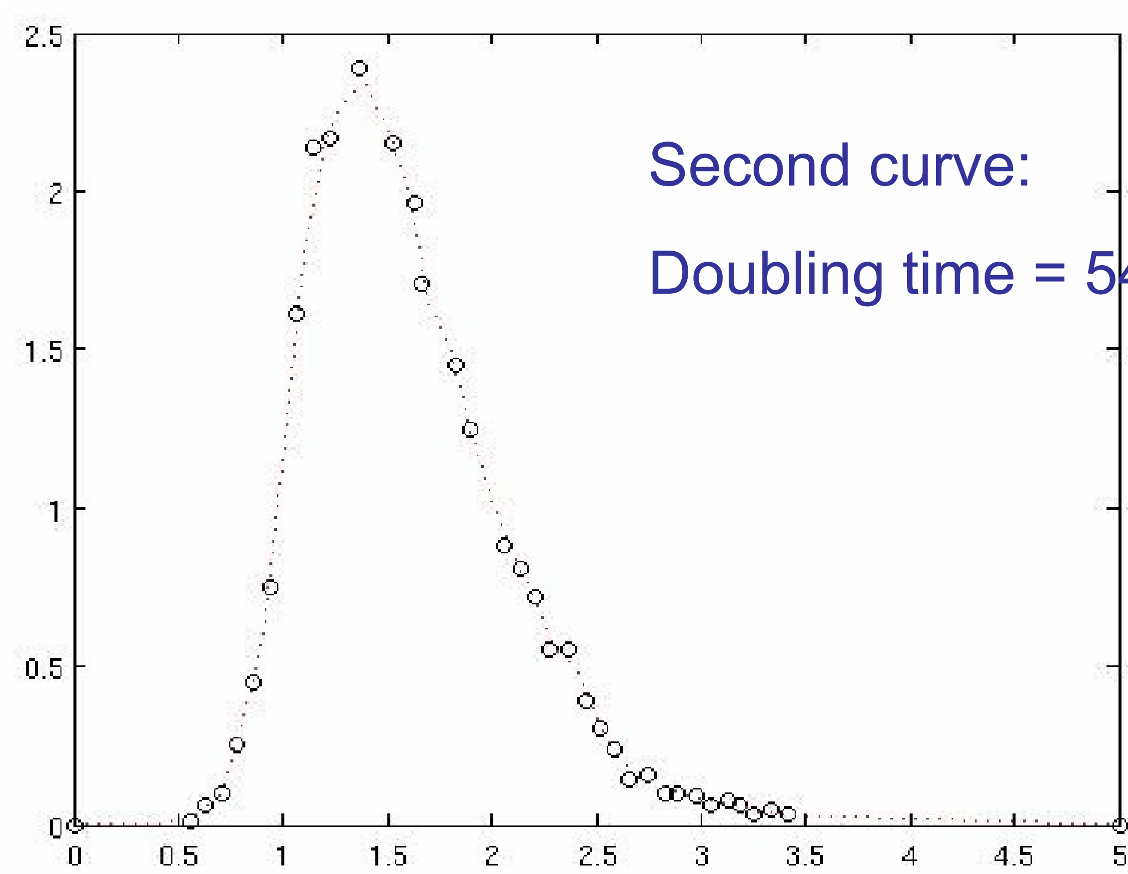
Application to Kubitschek's data

4 kinds of growth environment for E. Coli:

Second curve:

Doubling time = 54 mns

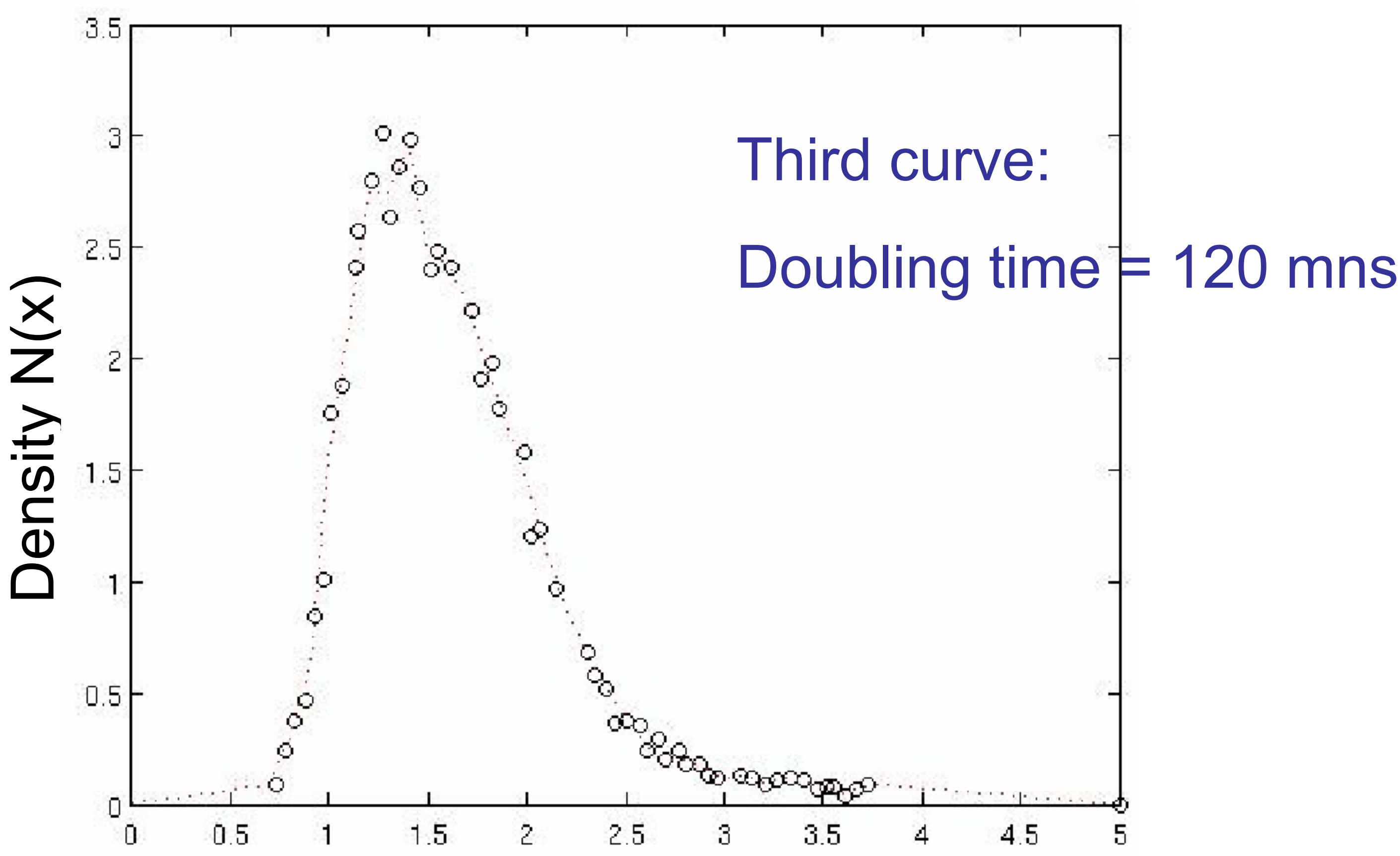
Density N(x)



Relative size (x=2: mean doubling size)

Application to Kubitschek's data

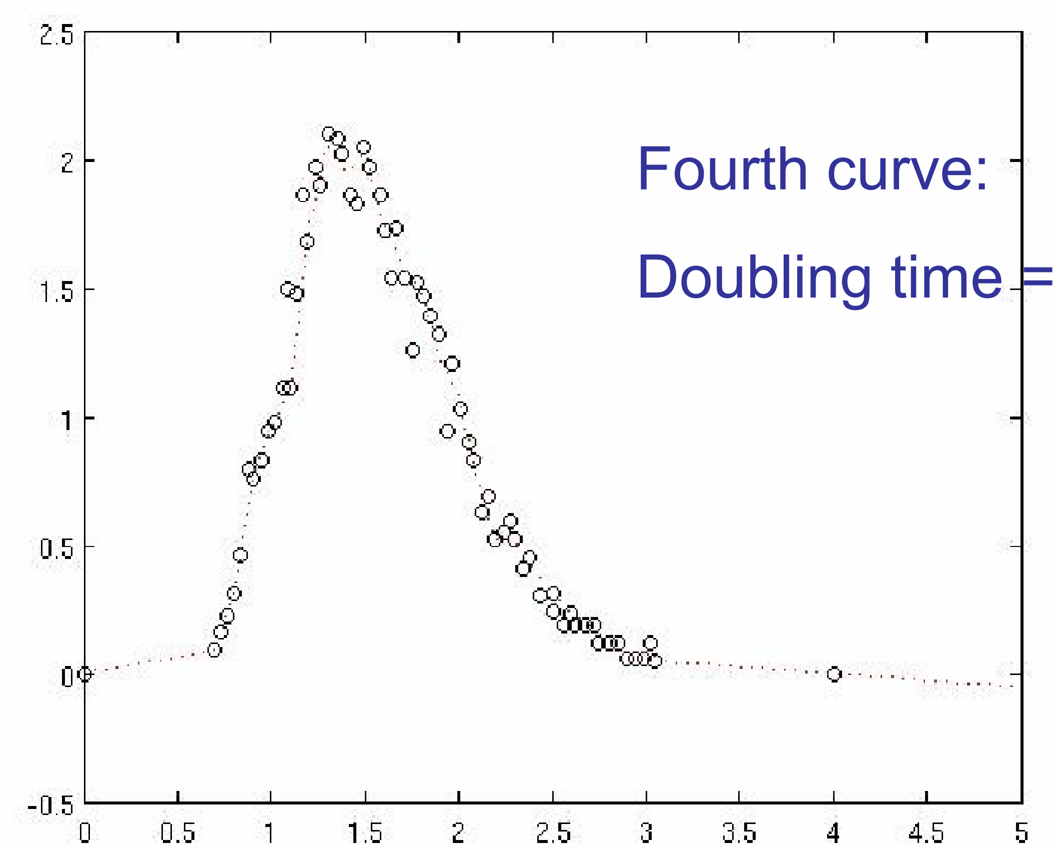
4 kinds of growth environment for E. Coli:



Relative size (x=2: mean doubling size)

Application to Kubitschek's data

4 kinds of growth environment for E. Coli:



Fourth curve:

Doubling time = 720 mns

Density N(x)

Relative size (x=2: mean doubling size)

Plan of work

- Transcribe the reported measurement and complete the boundary data for x close to 0 and large enough by 0
- Interpolate to a uniform grid (e.g. using splines)
- Deduce from the knowledge of both N. and λ_0 :
  - Doubling time $T_0 = \text{Log}(2) / \lambda_0$
  - Growth speed:
    - $g(x) = \lambda_0\, x$ (exponential growth)
    - $g(x)=\lambda_0 \int xN dx / \int N dx$ (linear growth)
- Performe a search for a good regularization parameter α
- Study the behavior of the solutions to the inverse problem by varying the regularization parameter.
- Study the consistency of the computed limiting distribution for the reconstructed B and the input data by computing the direct problem for some variants of the reconstructed B.

Application to Kubitschek's data

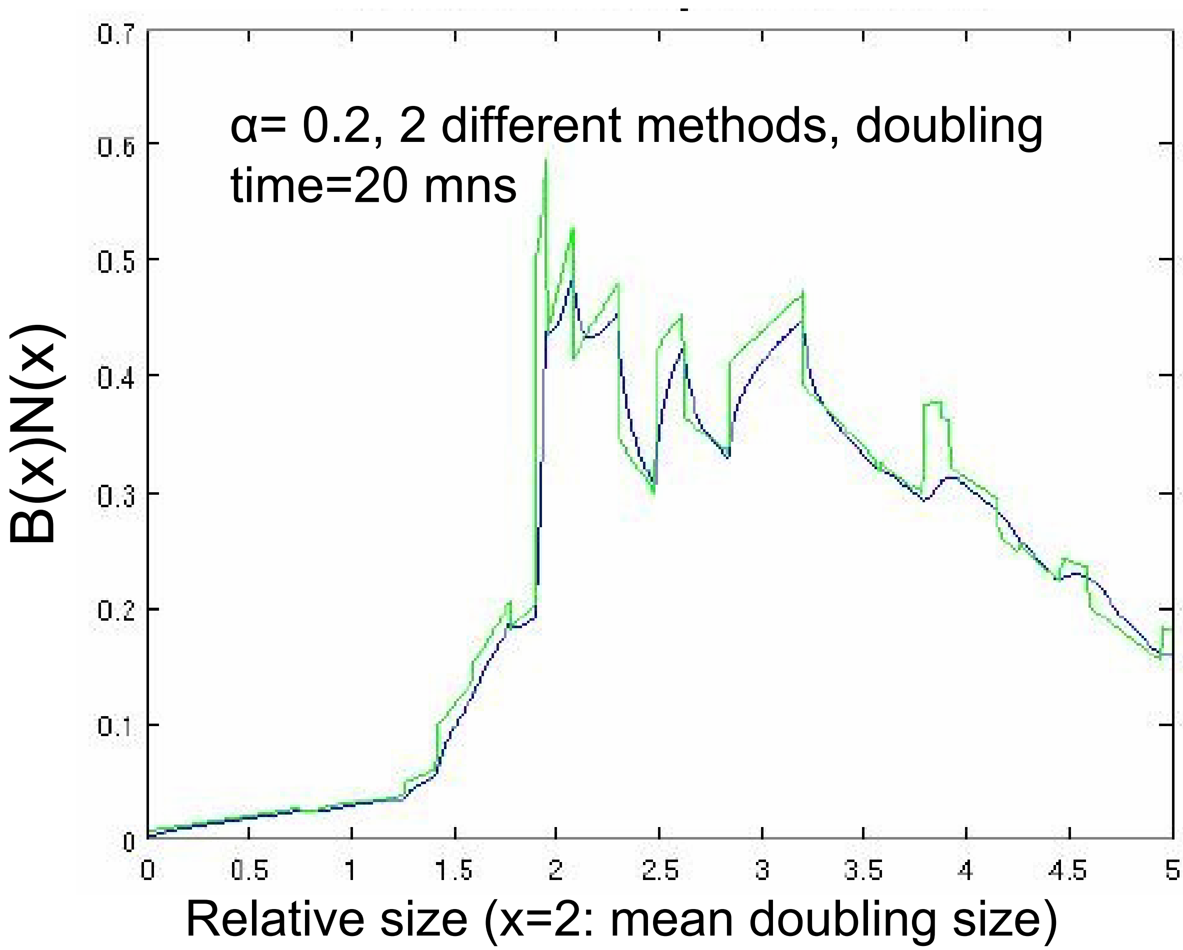
Assumption: Linear growth, constant growth speed calculated from the knowledge of the doubling time.

ratio := int(Nx)/int (N)

1) 1.8551
2) 1.5964
3) 1.6158
4) 1.5577

*g = ln(2)*ratio/doubling time*

1) 0.0643
2) 0.0205
3) 0.0093
4) 0.0015



Application to Kubitschek's data

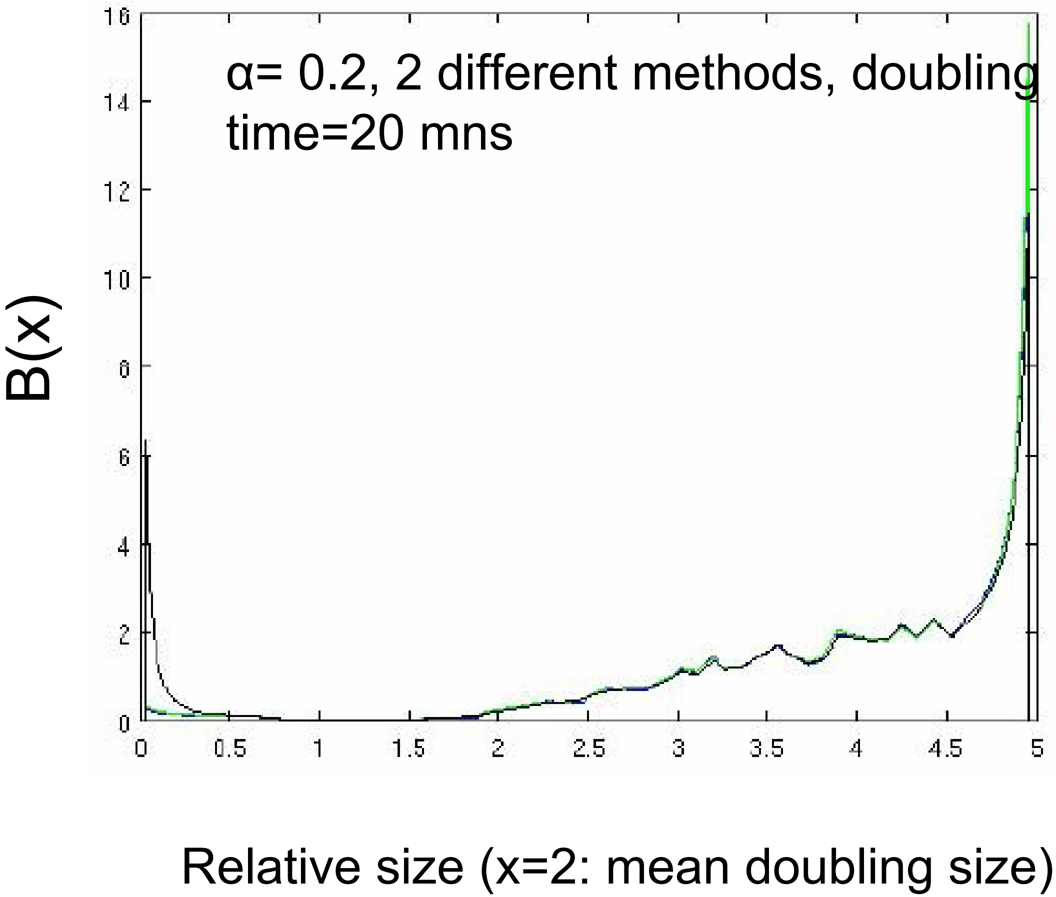
Assumption: Linear growth, constant growth speed calculated from the knowledge of the doubling time.

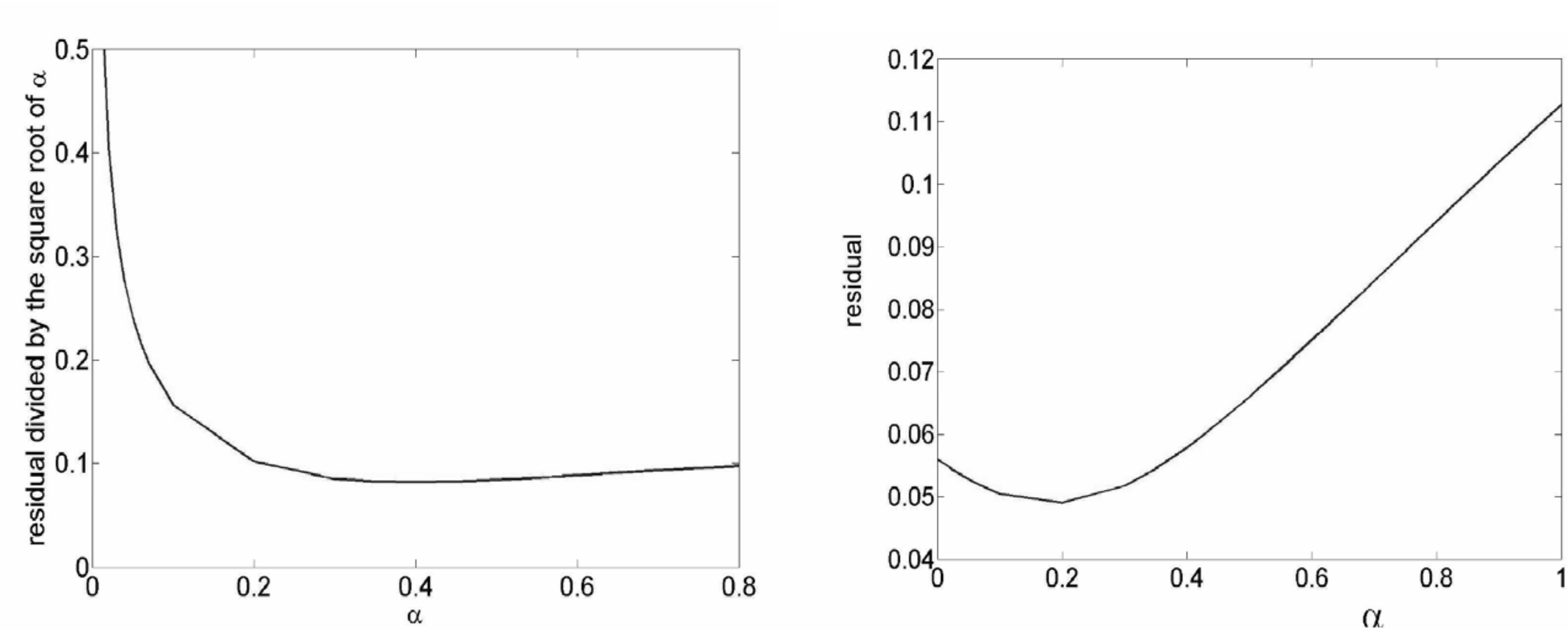
ratio := int(Nx)/int (N)

1) 1.8551
2) 1.5964
3) 1.6158
4) 1.5577

*g = ln(2)*ratio/doubling time*

1) 0.0643
2) 0.0205
3) 0.0093
4) 0.0015



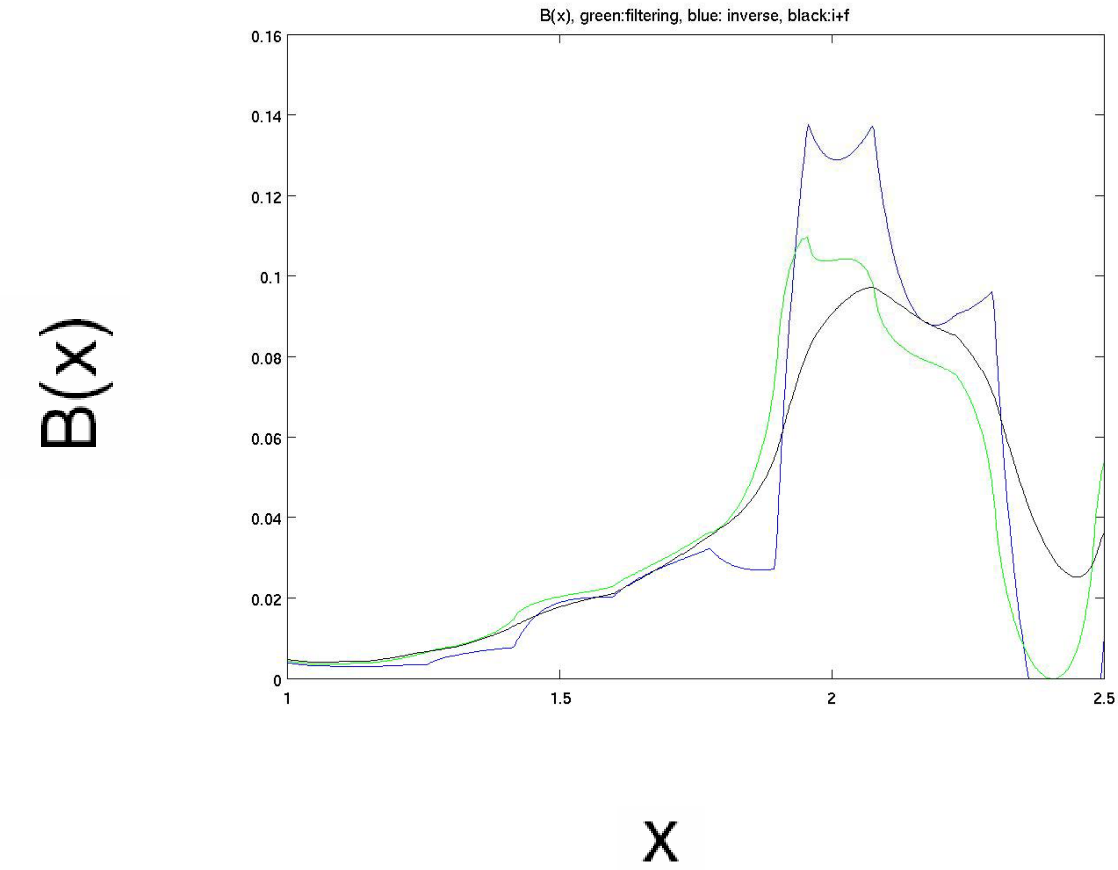


Example of criteria for the choice of α.

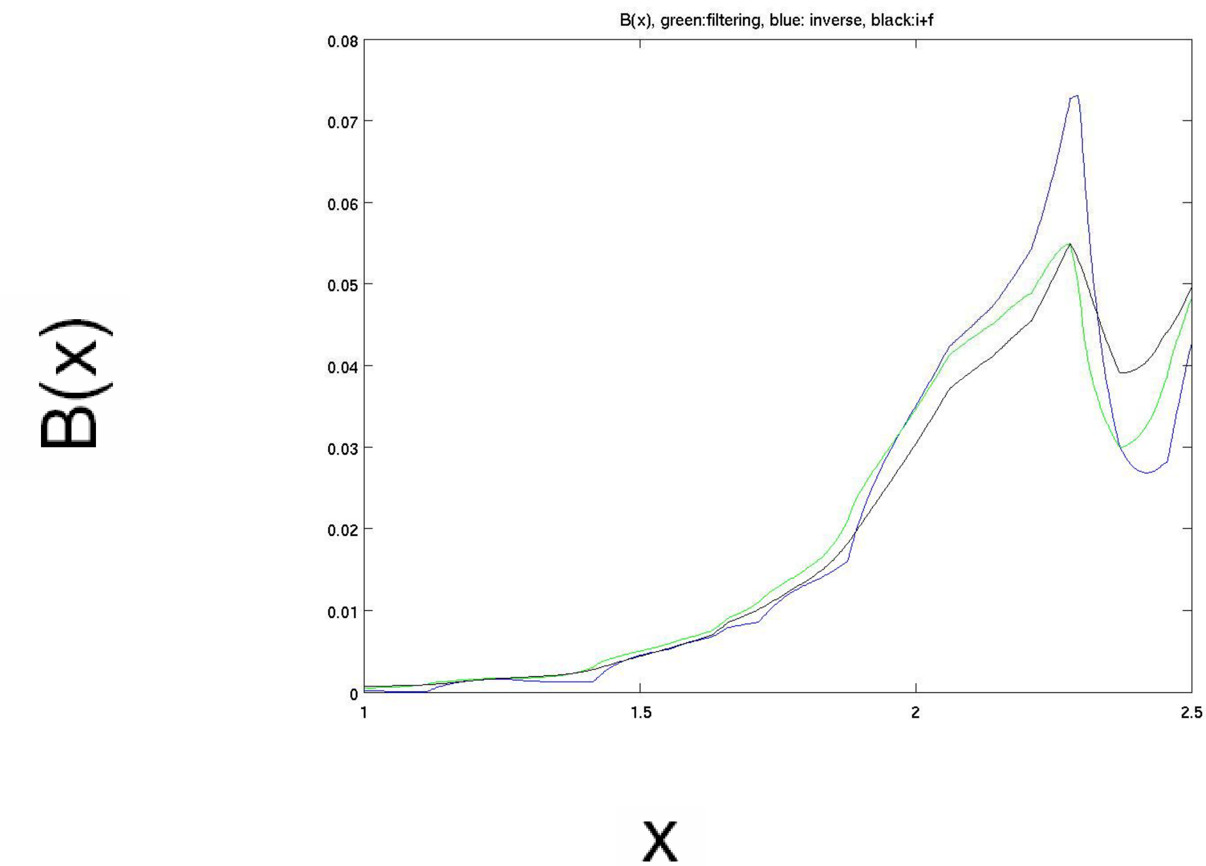
Left: for a doubling time of 20 min, curve of residual √α with respect to α, linear case. The minimum is attained for α = 0.4 but the curve is very flat for α = 0.2. (see Engl, Hanke, Neubauer, 1996)

Right: doubling time of 54 min and linear case, the curve of the residual exhibits a minimum for α = 0.2.

Doubling time: 20 minutes

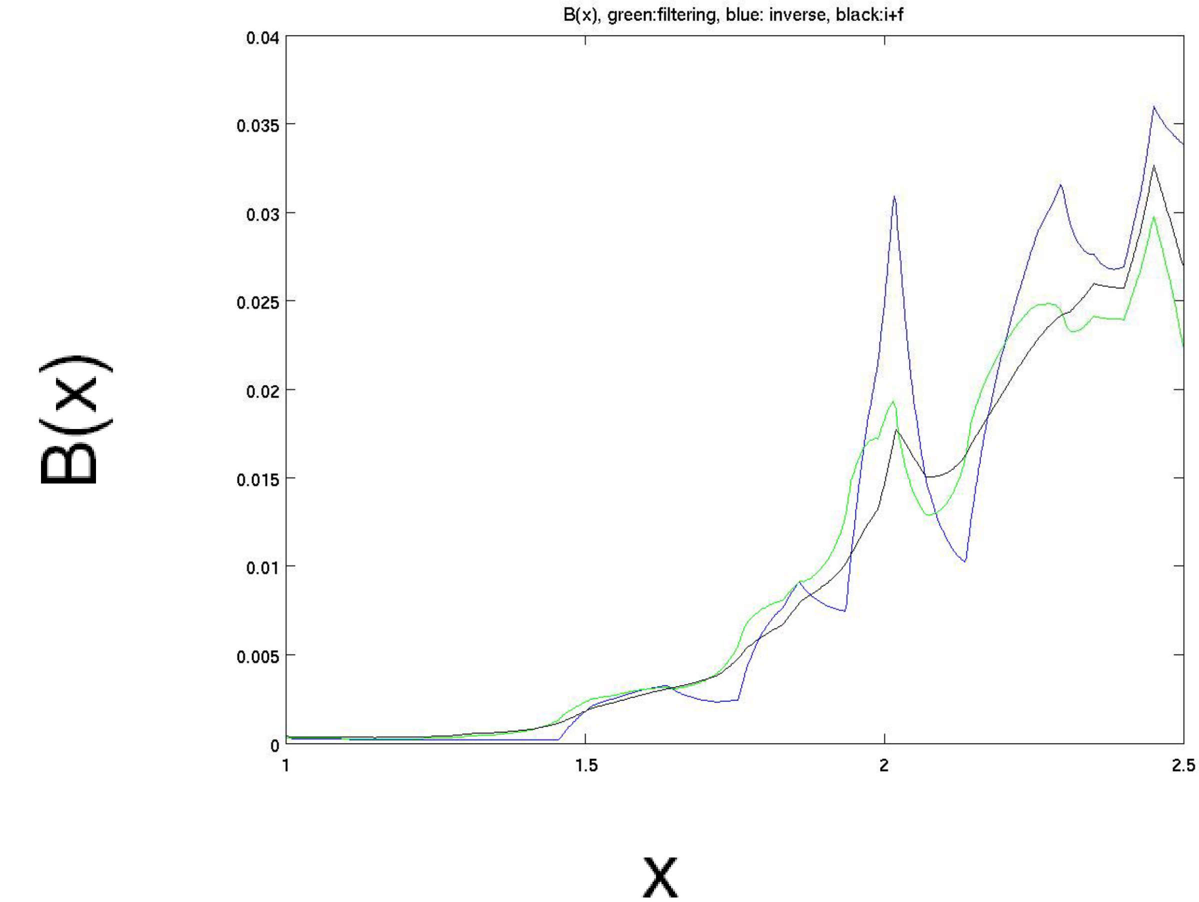


Doubling Time: 54 minutes

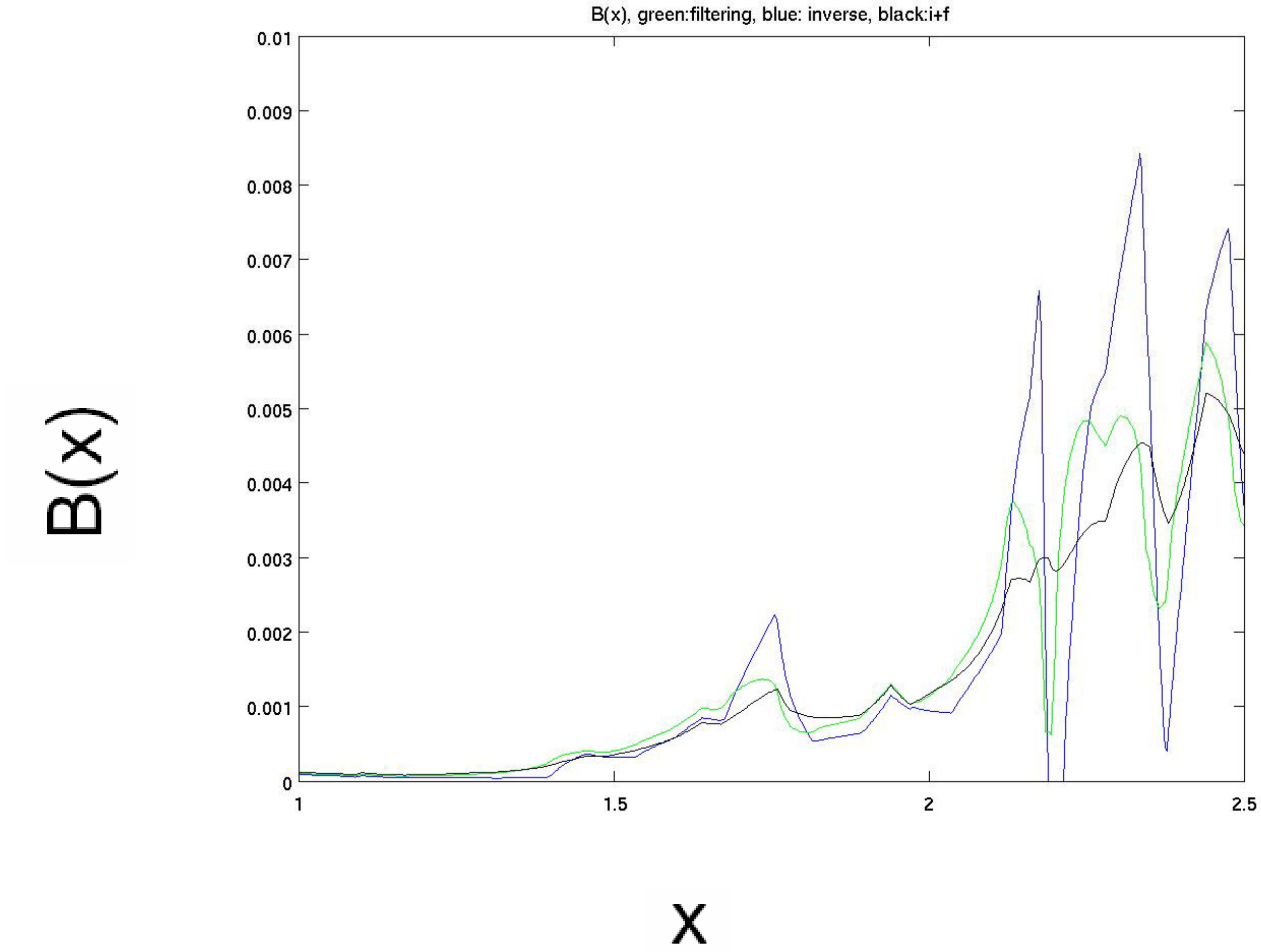


Marie DOUMIC (INRIA, BANG)
Nice, September 14th, 2009

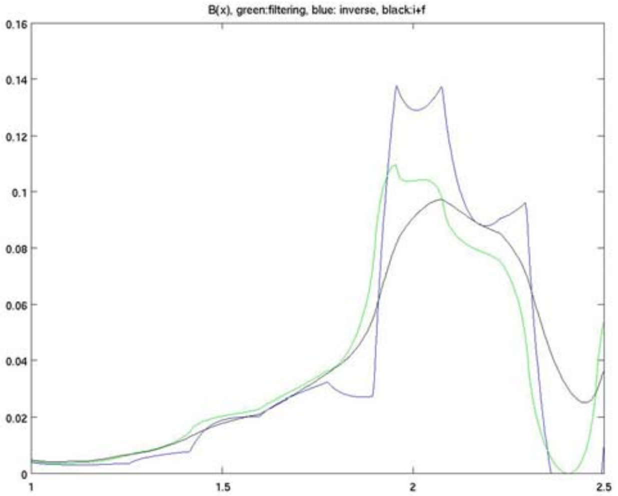
Doubling time: 120 minutes

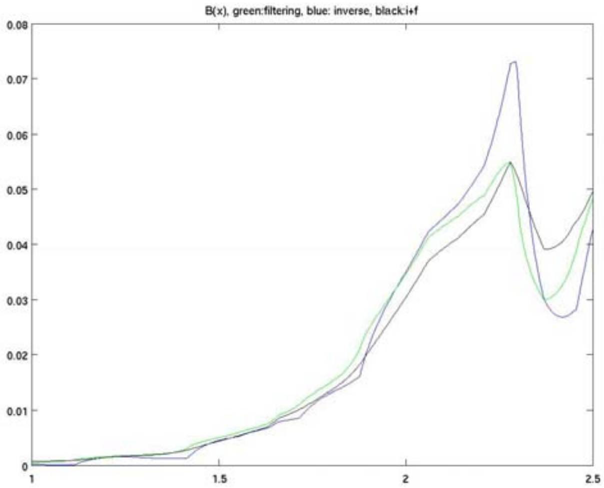


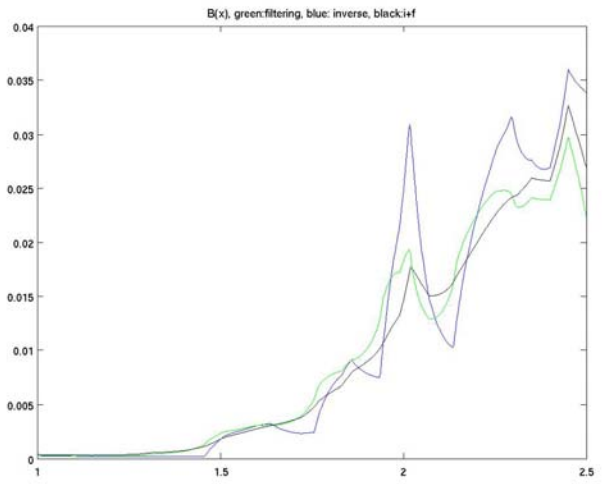
Doubling time: 720 minutes

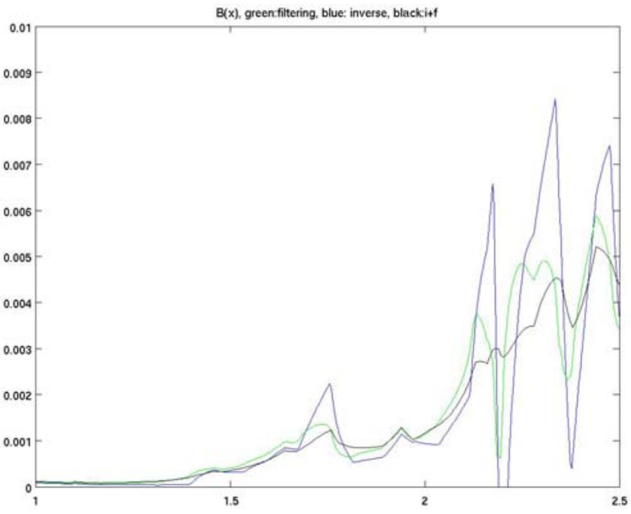


-> A birth pattern ?









Perspectives

- precise what is the noise
- Adaptation to recent data on plankton (M. Felipe)
- apply the method to adaptations of this model: especially for prion & Alzheimer diseases (protein polymerization processes – ANR grant TOPPAZ with H. REZAEI, I.N.R.A)

  → other fragmentation kernels

- improve the numerical method:
  - compare it with other regularizations like Tikhonov understand why the combination of both methods seems better
  - recover *B* globally, even where *N* vanishes, by getting *a priori* information on *B.*

C'est fini !