On the Efficient Identification of an Inflection Point

Demetris T. Christopoulos

***Abstract*—Our task is to find time efficient and statistically consistent estimators for revealing the true inflection point of a planar curve when we have only a probably noisy set of points for it, thus we are introducing extremum surface (ESE) and distance estimator (EDE) methods. The analysis is based on the geometric properties of the inflection point for a smooth function. Iterative versions of the methods are also given and tested. Numerical experiments are performed for the class of sigmoid curves and comparison with other available procedures is carried out. It is proven that both methods are quite fast in computational execution. Under a rather common noise type EDE can give a 96% confidence interval, while it always provides estimations for data with more than a million cases at a negligible execution time. An alternative way of mode computation for a distribution by using its CDF is given as a real massive data example.**

***Index Terms*—Inflection point identification, Efficient computation, Mode estimation, Iterarive methods.**

***MSC 2010 Codes* – 62F12, 65D17, 65D99**

I. Introduction

The sigmoid or S-shaped model is common in many disciplines like artificial neural networks [10], utility theory [7] & technological substitution [12], [14] in Economics, growth theory [2] & allosterism [1] in Biology, population dynamics [20], [18] in Ecology, autocatalysis [19] in Analytical Chemistry, dose-response [13] in Medicine and many others. A key concept for every sigmoid is the inflection point and its accurate and fast computation while the available methods can be classified to (i) adoption of a suitable *process* - test function - then use of regression or maximum likelihood estimation techniques, (ii) smoothing data under the restriction that *second order divided differences* change sign once [5], (iii) use of *Differential Geometry* to define a proper measure of discrete curvature and then searching for the least measure point [16], [9], Gaussian smoothing techniques [15] and shape analysis procedures [11].

At this paper we are presenting two *geometric methods* in order to find the inflection point of a curve, when we know only a set of points $\{(x_i, y_i)\,,\, i = 0, 1, \ldots, n\}$ for it, the *Extremum Surface Estimator* (*ESE* method) which works with planar surfaces and the *Extremum Distance Estimator* (*EDE* method) which uses only planar distances. We do not perform neither regression nor splines representation analysis. In addition, no concepts like discrete or digital curvature are defined or used. Instead we focus on finding each time two proper points where the true inflection point lies between and then we take as an estimator their middle point, just like *bisection method* works in root finding.

Demetris T. Christopoulos is with the Department of Economics, National and Kapodistrian University of Athens, Greece. (E-mail: dchristop@econ.uoa.gr)

Our main task is to evaluate a total of six methods, in order to find the most efficient one and use it to estimate the mode when we have massive data sets. *Paper structure*: 1. introduction, 2. presentation of *ESE* & *EDE* methods, 3. generalisation to their iterative versions, 4. evaluation and application limits, 5. an example of a real data set, 6. conclusion – proofs are given as Appendix.

II. Extremum Surface and Distance Estimator methods

A. Required concepts

Let a function $f : [a, b] \to R,\ f \in C^{(n)},\ n \geq 2$ which is convex for $x \in [a, p]$ and concave for $x \in [p, b]$, p is the unique inflection point of f in $[a, b]$ and let an arbitrary $x \in [a, b]$.

Definition 2.1: Total, left and right chord are the lines connecting points $\{(a, f(a)), (b, f(b))\}$, $\{(a, f(a)), (x, f(x))\}$ and $\{(x, f(x)), (b, f(b))\}$ with Cartesian equations $g(x)$, $l(x)$ and $r(x)$ respectively. Distances from chords are the functions $F, F_l, F_r : [a, b] \to R$ with

$$F(x) = f(x) - g(x), F_l(x) = f(x) - l(x), F_r(x) = f(x) - r(x) \tag{1}$$

By using elementary *Analytical Geometry* we can prove that

$$F\left(x\right) = f\left(x\right) - \frac{b\, f\left(a\right) - a\, f\left(b\right)}{b - a} - \frac{f\left(b\right) - f\left(a\right)}{b - a}\, x \tag{2}$$

$$F_l(t) = f\left(t\right) - f\left(a\right) + \frac{f\left(a\right) - f\left(x\right))}{x - a}\ \left(t - a\right)\ ,\ t \in [a, x] \tag{3}$$

$$F_r(t) = f\left(t\right) - f\left(x\right) + \frac{f\left(b\right) - f\left(x\right)}{b - x}\ \left(t - x\right)\ ,\ t \in [x, b] \tag{4}$$

Definition 2.2: The s-left $(s_l(a, x))$ and s-right $(s_r(b, x))$ are the algebraic surfaces

$$s_l(a, x) = \int_a^x F_l(t)dt \quad , \quad s_r(b, x) = \int_x^b F_r(t)dt \tag{5}$$

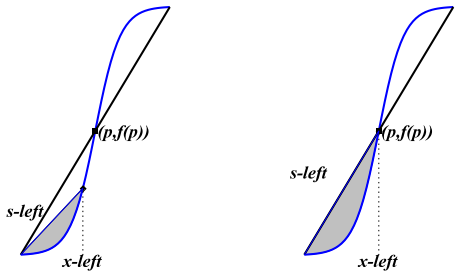
The x-left (x_l) and x-right (x_r) are x–values such that

$$x_l = \underset{x \in [a, b+\delta_1]}{argmin}\ \{s_l\left(a, x\right)\} \quad , \quad x_r = \underset{x \in [a-\delta_2, b]}{argmax}\ \{s_r\left(b, x\right)\} \tag{6}$$

with $\delta_1, \delta_2 > 0$ taken as small as necessary for x_l, x_r to be unique unconstrained extremes in the corresponding intervals. A graphical illustration of the above defined left and right-terms is presented at Fig. 1 and 2 where we observe that when $x = x_l$, $x = x_r$ then we achieve the algebraic minimum $s_l(a, x)$ and maximum $s_r(b, x)$, respectively. The motivation for the *left-* and *right-* naming was due to the origin of the chords: left/right starts from the left/right edge of the graph.

Definition 2.3: Normal algebraic distance of the curve point $(x, f(x))$ from the total chord Def. 2.1 is function $N : [a, b] \to R$ with

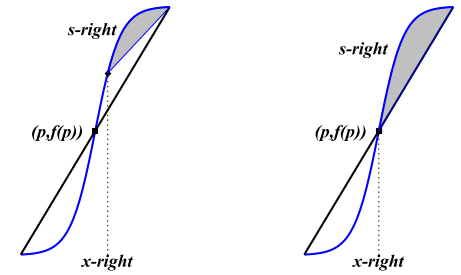
$$N(x) = -\frac{(f(b) - f(a))x - (b - a)f(x) + b\, f(a) - a\, f(b)}{\sqrt{(f(b) - f(a))^2 + (b - a)^2}} \tag{7}$$



(a) $x < p$ (b) $x = p$

(c) $x = x_l$ (d) $x > x_l$

Fig. 1: Illustration of the algebraic surfaces $s_l(a, x) < 0$ and x_l.



(a) $x > p$ (b) $x = p$

(c) $x = x_r$ (d) $x < x_r$

Fig. 2: Illustration of the algebraic surfaces $s_r(b, x) > 0$ and x_r.

For a convex/concave curve the above definition gives $N(x) < 0$ when $x < p$ and $N(x) > 0$ if $x > p$. For a not necessarily strictly sorted grid of $n+1$ not equally spaced points $\{x_i\}$ of Eq. 8 –*standard partition*– we add errors to the values of a known function f and create the noisy data set $\{(x_i, y_i)\}$ by the process of Eq. 9.

$$a = x_0 \leq x_1 \leq \ldots \leq x_n = b \tag{8}$$

$$y_i = f(x_i) + \epsilon_i,\ \epsilon_i \sim iid(0, \sigma^2) \tag{9}$$

Our analysis is applicable for every distribution with zero mean, for example the uniform $U(-r, r)$ or normal $N(0, \sigma^2)$. If error is distributed without a zero mean, then results are ambiguous.

Definition 2.4: For the noisy data Eq. 9 we define the discrete distances from total, left and right chord as the values

$$Y_i = y_i - g(x_i), Y_{li} = y_i - l(x_i), Y_{ri} = y_i - r(x_i),\ i = 0, \ldots, n \tag{10}$$

Definition 2.5: Function f, around its inflection point p, is called symmetric when

$$f(p + x) + f(p - x) - 2f(p) = 0,\ \forall x \in R \tag{11}$$

and is called locally (ϵ, δ) asymptotically symmetric when:

$$|f(p + x) + f(p - x) - 2f(p)| < \epsilon,\ \forall x \in (p-\delta, p+\delta) \tag{12}$$

Definition 2.6: The function $f : [a, b] \to R$, with respect to its inflection point p, has

1) Data symmetry if $p - b = a - p$
2) Data left asymmetry if $p - b < a - p$
3) Data right asymmetry if $p - b > a - p$

The $f : [a, b] \to R$ is characterised as totally symmetric, if it is symmetric around p and has also data symmetry.

Definition 2.7: For every subsequent $x_i < x_j$ the elementary trapezoidal estimation is

$$\int_{x_i}^{x_j} f(x)dx \approx T_{i,j}(f, x_i, x_j) = \frac{f(x_i) + f(x_j)}{2}(x_j - x_i) \tag{13}$$

And for every standard partition the total trapezoidal estimation is

$$\int_a^b f(x)dx \approx T_{n+1}(f, a, b) = \sum_{i=0}^{n-1} T_{i,i+1}(f, x_i, x_{i+1}) \tag{14}$$

B. The Extremum Surface Method

We can prove that

Lemma 2.1: The x-left (x_l), x-right (x_r) are abscissae such that left, right chord respectively are tangent to the graph $G(f)$.

Corollary 2.1: For Eq. 6 it holds

$$\begin{aligned} x_l &= \underset{x \in [a, b+\delta_1]}{arg\,x} \left\{ f'(x) = \tfrac{f(x)-f(a)}{x-a} \right\} \\ x_r &= \underset{x \in [a-\delta_2, b]}{arg\,x} \left\{ f'(x) = \tfrac{f(b)-f(x)}{b-x} \right\} \end{aligned} \tag{15}$$

with $\delta_1, \delta_2 > 0$ taken as small as necessary for x_l, x_r to be unique unconstrained solutions in the corresponding intervals. This tangency condition is illustrated at Fig. 1(c) and 2(c).

Corollary 2.2: Let a function $f : [a, b] \to R$, $f \in C^{(n)}$, $n \geq 2$ which is convex for $x \in [a, p]$ and concave for $x \in [p, b]$. Then we have one of the following possibilities:

1) If $x_l, x_r \in [a, b]$ then $a \leq x_r < x_l \leq b$
2) If $x_l \notin [a, b]$ then $x_l > b$
3) If $x_r \notin [a, b]$ then $x_r < a$

We define the next theoretical estimator of the inflection point:

Definition 2.8: The theoretical extremum surface estimator (TESE) is

$$x_S = \left\{ \begin{array}{ccc} \frac{x_l+x_r}{2} & , & x_l, x_r \in [a,b] \\ \frac{b+x_r}{2} & , & x_l > b \\ \frac{x_l+a}{2} & , & x_r < a \end{array} \right\} \quad (16)$$

Lemma 2.2: If the mesh $\lambda(n)$ of the standard partition is such that $\lim_{n\to\infty} n\lambda(n)^2 = 0$ then $T_{n+1}(y,a,b)$ is a consistent estimator of the $T_{n+1}(f,a,b)$.

Now we are able to compute using our trapezoidal rule of Def. 2.7 data estimations for $s_l(x_0,x_j)$ and $s_r(x_n,x_j)$:

Definition 2.9: The data estimators for the algebraic surfaces of Def. 5 are

$$\begin{array}{c} s_{l,j+1}(x_0,x_j) = T_{j+1}(Y_l,x_0,x_j) \\ s_{r,n-j+1}(x_n,x_j) = T_{n-j+1}(Y_r,x_j,x_n) \end{array} \quad (17)$$

Let us define data estimators for x_l, x_r.

Definition 2.10: The $\chi_l,\ \chi_r$ are values such that:

$$\chi_l = x_{j_l},\ j_l = \underset{j\in[1,n]}{argmin}\{s_{l,j+1}(x_0,x_j)\} \quad (18)$$

$$\chi_r = x_{j_r},\ j_r = \underset{j\in[0,n-1]}{argmax}\{s_{r,n-j+1}(x_n,x_j)\} \quad (19)$$

We define now the noisy data estimator of the inflection point:

Definition 2.11: The data extremum surface estimator (ESE) is

$$\chi_S = \frac{\chi_l + \chi_r}{2} \quad (20)$$

Lemma 2.3: The ESE is a consistent estimator of TESE with all relevant integrals calculated via trapezoidal rule.

We have to make a remark about the concept of convex or concave area, as is defined in [16] and as defined in this work. There the concept of area that is computed is a summation of the distances from a chord and curve, between two critical points, while our approach computes an actual geometric area by using the trapezoidal rule.

C. The Extremum Distance Method

Definition 2.12: The xF-left (x_{F1}), xF-right (x_{F2}) and xN-left (x_{N1}), xN-right (x_{N2}) are values such that:

$$x_{F1} = \underset{x\in[a-\delta_1,b]}{argmin}\{F(x)\},\quad x_{F2} = \underset{x\in[a,b+\delta_2]}{argmax}\{F(x)\} \quad (21)$$

$$x_{N1} = \underset{x\in[a-\delta_1,b]}{argmin}\{N(x)\},\quad x_{N2} = \underset{x\in[a,b+\delta_2]}{argmax}\{N(x)\} \quad (22)$$

with $\delta_1, \delta_2 > 0$ taken as small as necessary for x_{F1}, x_{F2} to be unique unconstrained extremums in the corresponding intervals.

The defined points and the corresponding line segments can be found at the Fig. 3 where it is also obviously shown that when we achieve a stationary value for $F(x)$ of Def. 2.1, then we achieve also the relevant stationary value for normal distance $N(x)$ of Def. 2.3, since the vector defined from $N(x)$ is just the orthogonal projection of the vector defined from $F(x)$ at the normal vector to the total chord. Now we shall prove the next useful Lemma.

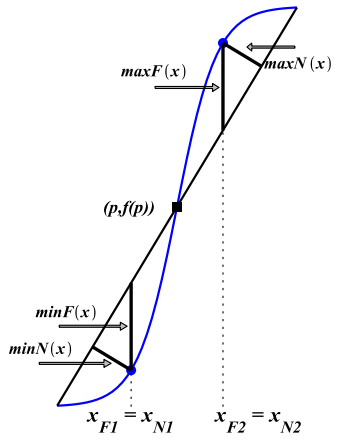


Fig. 3: Illustration of x_{F1}, x_{F2} and $minF(x)$, $maxF(x)$

Lemma 2.4: For the points of Eq. 21 it holds

$$x_{F1,2} = \underset{x\in[a-\delta_1,b+\delta_2]}{arg\,x}\left\{f'(x) = \frac{f(b)-f(a)}{b-a}\right\} \quad (23)$$

with $\delta_1, \delta_2 > 0$ taken as small as necessary for x_{F1}, x_{F2} to be unique unconstrained extrema in the corresponding intervals.

Corollary 2.3: Let a function $f : [a,b] \to R$, $f \in C^{(n)}$, $n \geq 2$ which is convex for $x \in [a,p]$ and concave for $x \in [p,b]$. Then we have one of the following possibilities:

1) If $x_{F1}, x_{F2} \in [a,b]$ then $a \leq x_{F1} < x_{F2} \leq b$
2) If $x_{F1} \notin [a,b]$ then $x_{F1} < a$
3) If $x_{F2} \notin [a,b]$ then $x_{F2} > b$

We define the next theoretical estimator of the inflection point:

Definition 2.13: The theoretical extremum distance from total chord estimator (TEDE) is such that

$$x_D = \frac{x_{F1} + x_{F2}}{2} \quad (24)$$

Let now define data estimators of x_{F1}, x_{F2} and x_D.

Definition 2.14: The data estimations of points defined at Eq. 21 are

$$\chi_{F1} = x_{j_1},\ j_1 = \underset{j\in[0,n]}{argmin}\{Y_j\} \quad (25)$$

$$\chi_{F2} = x_{j_2},\ j_2 = \underset{j\in[0,n]}{argmax}\{Y_j\} \quad (26)$$

Definition 2.15: The extremum distance from total chord estimator (EDE) is

$$\chi_D = \frac{\chi_{F1} + \chi_{F2}}{2} \text{ iff } \chi_{F2} \geq \chi_{F1} \quad (27)$$

Lemma 2.5: The EDE is an unbiased estimator of TEDE.

At this stage we have to mention that there exists a similar work with distances from the chord, see [9], where a summation of the relevant distances from the chord is taken in order to define the concept of a *discrete curvature* for a planar curve. An analogous approach is that of [16], where it is also used a proper summation of the distances between the chord and the curve points. Here we do not define and we do not compute any kind of curvature, but we just choose only the two extreme

distances needed for Def. 2.12. Another interesting property of EDE is that we can give a 96% *Chebyshev* confidence interval because we can find an estimation of the error variance using next

Lemma 2.6: Let a data set created from a known function f by using a strictly zig–zag process $y_i = f(x_i) + (-1)^i \epsilon_i$ for adding *iid* error terms $\epsilon_i \sim U(0, r)$. Then it holds that

$$s^2 = Var(\widehat{\boldsymbol{y} - \mathbf{f}}(\mathbf{x})) \approx \frac{2}{n} \sum_{i=1}^{n} \left(\frac{y_i - y_{i-1}}{2} \right)^2 \tag{28}$$

It is easy to prove also the next

Lemma 2.7: The variance of *EDE* method which is applied for a data set created using the requirements of Lemma 2.6 is

$$s_D^2 = \frac{1}{2} s^2 = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - y_{i-1}}{2} \right)^2 \tag{29}$$

Corollary 2.4: The 96% *Chebyshev* confidence interval for EDE estimator of data under requirements of Lemma 2.6 is

$$\left[\chi_D - 5 \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - y_{i-1}}{2} \right)^2}, \chi_D + 5 \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i - y_{i-1}}{2} \right)^2} \right] \tag{30}$$

III. Iterative application of ESE & EDE methods

Another important option is the possibility of iterations similar to those of the well known *bisection method* in root finding. Recall that for a continuous function if $f(\alpha)\, f(\beta) < 0$ then exists $\xi \in (\alpha, \beta)$ such that $f(\xi) = 0$, while its first estimation is $\chi^{(0)} = \frac{a+b}{2}$. Our *ESE* method always gives an interval that contains the true inflection point p or a point close to the edge a or b, if data is just convex (or just concave) and inflection point does not exist. The *EDE* method also gives an interval in most cases, although it is more sensitive to errors of x_0, x_n, so it does not always give a point close to a or b, if simple convexity or concavity exist.

ESE iterative method or Bisection-ESE (BESE)

We apply to the probably noisy data $\{(x_i, y_i), i = 0, \ldots, n\}$ our *ESE* method and obtain the 0^{th} estimation

$$[j_r^{(0)}, j_l^{(0)}],\ \chi_r^{(0)} = x_{j_r^{(0)}},\ \chi_l^{(0)} = x_{j_l^{(0)}},\ \chi_S^{(0)} = \frac{\chi_r^{(0)} + \chi_l^{(0)}}{2} \tag{31}$$

If and only if $j_l^{(0)} > j_r^{(0)}$, then we apply again *ESE* for data $\left\{(x_i, y_i), i = j_r^{(0)}, \ldots, j_l^{(0)}\right\}$ and obtain the 1^{st} estimation

$$[j_r^{(1)}, j_l^{(1)}],\ \chi_r^{(1)} = x_{j_r^{(1)}},\ \chi_l^{(1)} = x_{j_l^{(1)}},\ \chi_S^{(1)} = \frac{\chi_r^{(1)} + \chi_l^{(1)}}{2} \tag{32}$$

We continue until $j_l^{(k)} < j_r^{(k)}$ or until the number of data pairs becomes five, since it is meaningless for the method to proceed with less than five points.

EDE iterative method or Bisection-EDE (BEDE)

We use for initial data $\{(x_i, y_i), i = 0, \ldots, n\}$ the *EDE* method and find the 0^{th} estimation *iff* $\chi_{F2} > \chi_{F1}$

$$[j_1^{(0)}, j_2^{(0)}],\ \chi_{F1}^{(0)} = x_{j_1^{(0)}},\ \chi_{F2}^{(0)} = x_{j_2^{(0)}},\ \chi_D^{(0)} = \frac{\chi_{F1}^{(0)} + \chi_{F2}^{(0)}}{2} \tag{33}$$

If and only if $j_2^{(0)} > j_1^{(0)}$, then we use again *EDE* method for data $\left\{(x_i, y_i), i = j_1^{(0)}, \ldots, j_2^{(0)}\right\}$ and find the 1^{st} estimation, again *iff* $\chi_{F2}^{(2)} > \chi_{F2}^{(1)}$

$$[j_1^{(1)}, j_2^{(1)}],\ \chi_{F1}^{(1)} = x_{j_1^{(1)}},\ \chi_{F2}^{(1)} = x_{j_2^{(1)}},\ \chi_D^{(1)} = \frac{\chi_{F1}^{(1)} + \chi_{F2}^{(1)}}{2} \tag{34}$$

while we use the same stopping criterion as in *BESE*. In both *BESE* & *BEDE* methods we begin from the initial interval $[a_0, b_0] = [a, b] = [x_0, x_n]$ and find the 0^{th} estimation $\chi_S^{(0)}$, $\chi_D^{(0)} \in [a_0, b_0]$, then we continue with a second interval $[a_1, b_1] \subset [a_0, b_0]$– which is either $[a_1, b_1] = [\chi_r^{(0)}, \chi_l^{(0)}]$ (*BESE*) or $[a_1, b_1] = [\chi_{F1}^{(0)}, \chi_{F2}^{(0)}]$ (*BEDE*)– and compute the 1^{st} estimation $\chi_S^{(1)}$, $\chi_D^{(1)} \in [a_1, b_1]$ and so on until the termination criteria are applicable. So, in essence, the only difference from *bisection method* is that now edges of interval $[a_{i+1}, b_{i+1}]$, which is the next step $i+1$, are not taken by the requirement $f(a_{i+1}) f(b_{i+1}) < 0$ but are the $[\chi_r^{(i)}, \chi_l^{(i)}]$ or $[\chi_{F1}^{(0)}, \chi_{F2}^{(0)}]$ outputs of the previous step i, see Fig. 4 for *BESE* iterations of a *Gompertz* noisy data, created using $f(x) = 1000\, \mathrm{e}^{-\mathrm{e}^{5} \mathrm{e}^{-\frac{x}{100}}}$ starting from $[a_0, b_0] = [347, 960]$ and adding error term $\epsilon_i \sim U(-25, +25)$ at an equal spaced x–grid with $n = 4086$. For the same function but for $[a_0, b_0] = [0, 1000]$ and an

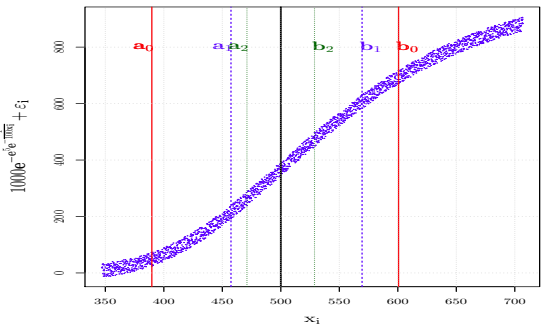


Fig. 4: Bisection ESE (BESE) iterations for noisy data

equidistant grid with $n = 10^7$ we apply $BEDE$ without error term, just to show its fast convergence to the true inflection point, see Table I and Fig. 5, where we show labels only for the first four iterations. It is obvious that,

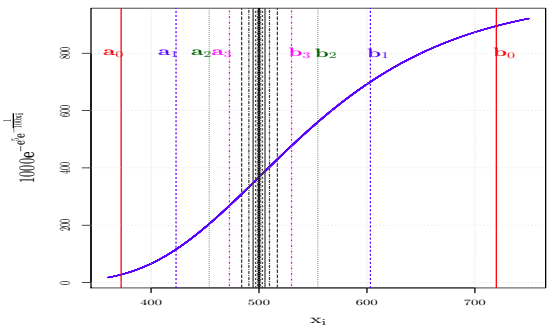


Fig. 5: Bisection EDE (BEDE) iterations for accurate data

although *Gompertz* sigmoid curve is not either symmetric

TABLE I: First five BEDE iterations form a total of 21

i	$\chi_{F1}^{(i)}$	$\chi_{F2}^{(i)}$	$\chi_{D}^{(i)}$	$N=n+1$
0	372.2823	719.8335	546.0579	10000001
1	423.1383	603.1799	513.1591	3475513
2	453.8448	554.5334	504.1891	1800417
3	472.6048	530.1469	501.3759	1006887
4	483.9011	517.0116	500.4564	575422
5	490.6044	509.6994	500.1519	331106

around inflection point, thus our *EDE* method will not give an initial accurate value, the *BEDE* convergences to the true value after just three iterations. The number of points that will be used at next step follow approximately the binary exponential decay formula $N_k = 0.927\frac{N_0}{2^k}, k = 0, 1, \ldots,$, so it actually halved the number of points, after some initial steps. The needed time is always negligible, for example the totally elapsed time for all above computations was $1.58\,s$. Just for a comparison, if we try to use NLS non linear regression which uses $SSlogis(x, Asym, xmid, scal)\{stats\}$ method of [17] to find the inflection point of our initial 10^7+1 points we shall fail to obtain an answer, even if we use the $SSgompertz(x, Asym, b2, b3)\{stats\}$ method, since we observe after $123.87\,s$ that we cannot proceed further due to memory restrictions of R. So, even if we know the exact underlying model, we are not able to find its parameters when we have a very large number of observations.

IV. Evaluation and Application Limits

We shall use 6 methods in order to find each time the inflection point, our newly introduced ESE, $BESE$, EDE, $BEDE$ geometrical methods, the $L2CXCV$ method presented at [5] and the NLS non linear regression which uses $SSlogis\{stats\}$ method of *R*. All methods except last one have been implemented using double precision *FORTRAN*. Execution is done inside *R* environment by creating proper *dll's* using typical Intel Core i5 CPU with 4 GB RAM and a 32 bit OS. We design our numerical experiments by using two suitable sigmoid functions of known inflection point $p = 500$, an interval $[a, b]$ which is every time the $[1\%, 99\%]$ of their total capacity and we add a uniform error $\epsilon_i \sim U(-r, r)$ via the process Eq. 8 & 9. Our set of test functions has members the *Fisher-Pry* sigmoid with total symmetry, taken from [6], with use in *technological substitution models* $f_1(x) = 500 + 500\tanh\left(\frac{x}{100} - 5\right)$ for $x \in [a, b] = [270, 730]$ and the non symmetric *Gompertz* sigmoid, after [8], which is often used in *population dynamics* $f_2(x) = 1000\,\mathrm{e}^{-\mathrm{e}^{5}\mathrm{e}^{-\frac{x}{100}}}$ for $x \in [a, b] = [347, 960]$. The run times in milliseconds and the results for 1000 experiments are presented at Fig. 6, 8 and 7, 9, respectively for the two curves. We have also computed the 96% *Chebyshev* confidence interval for our *EDE Fisher-Pry* estimations, see Fig. 10, where we observe that the intervals' average width is approximately 4% of inflection point value.

Clearly the most time efficient method is *EDE* and its iterative version *BEDE*, while the *L2CXCV* method needs more time and has also many *outliers*. As for the accuracy, we have to remind that our totally symmetric *Fisher-Pry* sigmoid has *TESE* & *TEDE*, see Eq. 2.8 & 2.13 equal to 500, thus our methods give the theoretically predicted results. The

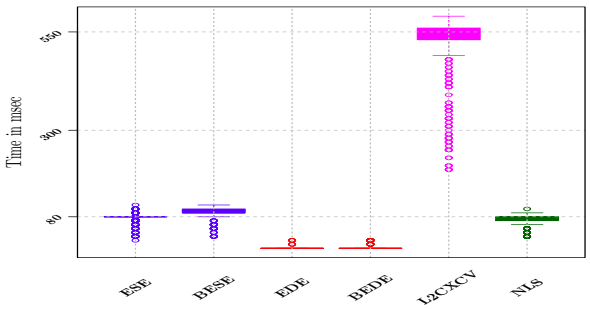


Fig. 6: Run times of all methods for 1000 Fisher-Pry sigmoid noisy data sets with N=1381 points

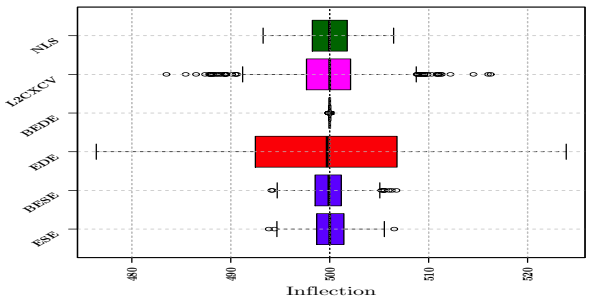


Fig. 7: Inflection point estimations for data of Fig. 6

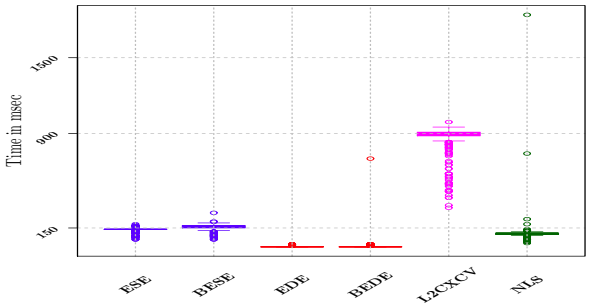


Fig. 8: Run times of all methods for 1000 Gompertz sigmoid noisy data sets with N=1840 points

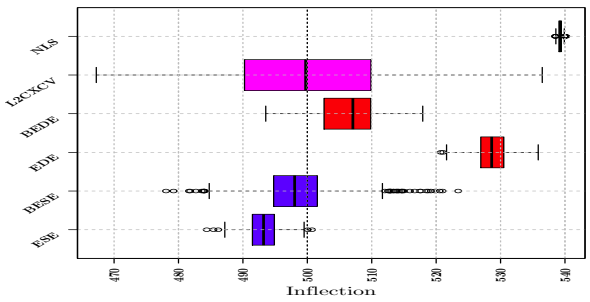


Fig. 9: Inflection point estimations for data of Fig. 8

same is true for the non symmetric *Gompertz* sigmoid, since $(TESE, TEDE) = (492.91, 528.77)$ and from either Fig. 9 or by computation we find their medians to be 493.21, 528.62 respectively, very close to theoretical values. So *BESE, BEDE* methods give medians close to exact value $p = 500$, while *L2CXCV* is perfectly symmetric around it. Only the *NLS* method is totally away, since we use as default test function the *SSLogis*. We next study the time needed for all our newly presented methods, in comparison to the previously existed, as a function of the number N of (x_i, y_i) pairs. For this task we create using Eq. 8 & 9 with $a = 270$, $b = 730$, $\epsilon_i \sim U(-10, +10)$ the Fisher-Pry sigmoid with total symmetry, see Table II, where we observe that EDE, $BEDE$ times are negligible, while ESE, $BESE$, $L2CXCV$ are not. Observe

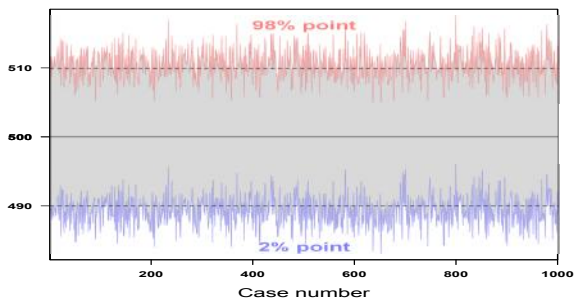


Fig. 10: 96% *Chebyshev* confidence interval for EDE at Fisher-Pry sigmoid

TABLE II: Execution times (in secs) with respect to N

N	EDE	BEDE	NLS	ESE	BESE	L2CXCV
10001	0.00	0.00	0.44	3.69	4.43	11.15
30001	0.00	0.00	1.42	32.83	37.31	195.49
50001	0.00	0.00	2.36	101.99	120.77	596.87
70001	0.00	0.03	1.94	183.96	178.59	925.64
100001	0.00	0.01	2.80	342.20	311.20	1756.07
$10^7 + 1$	0.53	1.07	–	–	–	–

here that for noisy data with 10 million xy–pairs we needed just 0.53/1.07 sec using *ESE, BEDE*, while other methods were practically non applicable.

V. A BIG DATA SET FROM REAL LIFE: FINDING THE MODE OF ANNUAL AVERAGE ATMOSPHERE TEMPERATURE FOR 214 YEARS

The inflection point of a *Cumulative DF* (*CDF*) corresponds to the maximum of *Probability DF* (PDF), since its second derivative is the first one of latter. An advantage of using *CDF*'s for big data comes from smoothing performed upon their estimation, since cumulation process absorbs local disturbances. We are using data that have been collected and processed in [4]. Our task is to find the main *mode* of average values for annual absolute temperature in *Northern Hemisphere* from $N = 388335$ rows of 1800–2013 yearly meteorological station observations. Their five point summary can be found at Table III.

TABLE III: NH annual average absolute temperature 1800–2013 description statistics

	Min	Q_1	Median	Mean	Q_3	Max
$^\circ C$	-29.050	4.950	9.192	9.667	14.520	32.100

Our first approach will be based on *Empirical CDF* and application of *BEDE* method, see Table IV, where after four iterations and just 0.08 sec we find $M \approx 9.279^\circ C$. If we use *NLS/Logistic* we find after 43.73 sec that a 95% confidence interval is $(8.9894^\circ C, 8.9928^\circ C)$. Other methods are practically unavailable for the original data set.

In order to apply them we start from the 13263 unique temperatures, compute first order differences for them and keep only values that have an absolute difference less than than $\approx 10^{-13}$ (the *FORTRAN* '*epsilon of the machine*'). This is required for *L2CXCV* method because it works only with strict sorted abscissae. Now as an estimation for the unknown *CDF* we are creating a cumulative frequency table with those 5968 temperature values as bins and use all methods presented at this paper to produce Table V and Fig. 11, where we have also estimated *PDF* using function *bkde* $\{$*KernSmooth*$\}$ [17], [22] with *Gaussian Kernel*. For a better visual comparison with *CDF* we have scaled *PDF* as shown in Fig. 11. Now the 95% confidence interval for *NLS/Logistic* is estimated to be $(9.2343^\circ C, 9.2681^\circ C)$, quite different and closer to other methods.

TABLE IV: BEDE estimations using ECDF of 388335 temperatures

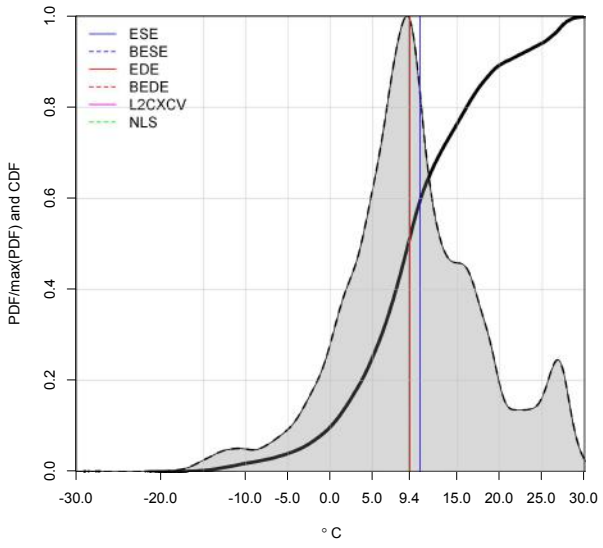
i	$\chi_{E1}^{(i)}$	$\chi_{E2}^{(i)}$	$\chi_{D}^{(i)}$	N
0	-0.691667	19.483333	9.395833	388335
1	4.450000	12.316667	8.383333	310480
2	6.750000	10.683333	8.716667	173517
3	7.841667	10.141667	8.991667	101947
4	8.741667	9.816667	9.279167	63023

TABLE V: mode of annual absolute temperature for years 1800–2013 in Northern Hemisphere

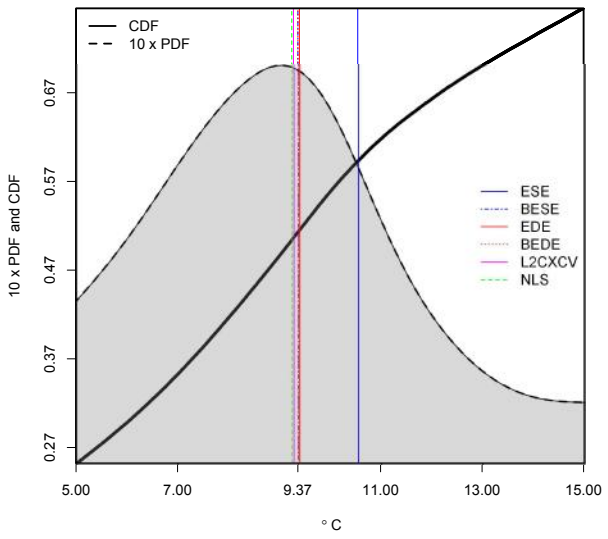
	ESE	BESE	EDE	BEDE	L2CXCV	NLS
mode ($^\circ C$)	10.550	9.371	9.400	9.371	9.283	9.251
Time (sec)	1.460	1.640	0.000	0.000	12.850	0.410

VI. CONCLUSION

Previously available methods can be applied for relatively small data sets, since their complexity do not allow them to always give an output in a short time. For example $L2CXCV$ requires strictly sorted abscissae and practically works until $n \approx 50000$. The model dependent NLS many times is not able even to give a first output if the underlying f is quite different than the adopted model for it. Provided that an inflection point is present inside our noisy data set, then Extremum Surface (*ESE*) and Distance (*EDE*) estimators can be used and give us a value close to their theoretical expected *TESE* (Def. 2.8) or *TEDE* (Def. 2.8). If data follow a *zig-zag* pattern, then a 96% *Chebyshev* confidence interval (Eq. 30) is available for EDE method with relative small width. If the underlying function is



(a) All available temperature data



(b) Only for interval $[5^\circ\, C, 15^\circ\, C]$

Fig. 11: Estimations of simple, cumulative density functions and mode, data as for Tables III and V

not symmetric around inflection point (Def. 11) or if we have data asymmetry (Def. 2.6) then we can use one of the iterative versions *BESE* & *BEDE* in order to move to a neighbourhood of p where f is at least locally asymptotically symmetric (Def. 12) and be able to accurately proceed. Numerical computations show that *EDE/BEDE* are the most time efficient and can give reliable output for data with more than a million rows in just a second. The *R Package inflection* is available [3] for using *ESE*, *BESE* & *EDE*, *BEDE* methods with *R*, while implementations in *FORTRAN*, *Maple* and *MATLAB* are also available. Applications to real data are easily performed and inflection point can always be found, provided that exists.

APPENDIX: PROOFS

Lemma 2.1 By using Eq. 3 & 5, smoothness conditions and the *Leibniz* rule for integral differentiation we find that $s_l^{'}(a,x) = \int_a^x \left(-\frac{f^{'}(x)}{x-a} - \frac{(f(a)-f(x))}{(x-a)^2} \right)(t-a)\,dt = -\frac{(x-a)\,f^{'}(x)-f(x)+f(a)}{2} = 0 \rightarrow f^{'}(x) = \frac{f(x)-f(a)}{x-a} = l^{'}(x)$. Now, using same rule, we find $s_l^{''}(a,x) = -\frac{1}{2}(x-a)\,f^{''}(x) \rightarrow s_l^{''}(a,x_l) > 0$, since $x_l > p > a$ and f is concave now. (Similar is the x_r–proof, using Eq. 4). □

Lemma 2.2 For every subsequent $x_i < x_{i+1}$ the elementary trapezoidal estimation is $T_{i,i+1}(y,x_i,x_{i+1}) = \frac{x_{i+1}-x_i}{2}y_i + \frac{x_{i+1}-x_i}{2}y_{i+1}$ and by taking the expected value we find that it holds $E\left(T_{i,i+1}(y,x_i,x_{i+1})\right) = \frac{x_{i+1}-x_i}{2}f(x_i) + \frac{x_{i+1}-x_i}{2}f(x_{i+1}) = T_{i,i+1}(f,x_i,x_{i+1})$, so from the linearity of expected value we have also that $E\left(T_{n+1}(y,a,b)\right) = \sum_{i=0}^{n-1} E\left(T_{i,i+1}(y,x_i,x_{i+1})\right) = T_{n+1}(f,a,b)$. Thus our estimator is unbiased. We continue by computing the variance of the elementary trapezoidal estimation $V\left(T_{i,i+1}(y,x_i,x_{i+1})\right) = \left(\frac{x_{i+1}-x_i}{2}\right)^2 V(y_i) + \left(\frac{x_{i+1}-x_i}{2}\right)^2 V(y_{i+1}) = \frac{(x_{i+1}-x_i)^2}{4}\sigma^2 + \frac{(x_{i+1}-x_i)^2}{4}\sigma^2 = \frac{(x_{i+1}-x_i)^2}{2}\sigma^2$ We have two cases. If standard partition is equal spaced, then $x_{i+1} - x_i = \frac{b-a}{n}$ and we obtain $V\left(T_{i,i+1}(y,x_i,x_{i+1})\right) = \frac{(b-a)^2}{2n^2}\sigma^2$. Let' s compute now the variance of estimator $T_{n+1}(y,a,b)$ and find $V\left(T_{n+1}(y,a,b)\right) = V\left(\sum_{i=0}^{n-1} T_{i,i+1}(y,x_i,x_{i+1})\right) = n\,V\left(T_{i,i+1}(y,x_i,x_{i+1})\right) = n\,\frac{(b-a)^2}{2n^2}\sigma^2 = \frac{(b-a)^2}{2n}\sigma^2$, thus it holds $\lim_{n\rightarrow\infty} V\left(T_{n+1}(y,a,b)\right) = \lim_{n\rightarrow\infty} \frac{(b-a)^2}{2n}\sigma^2 = 0$. For the second case, if standard partition is not equal spaced then the mesh or norm of the partition is $\lambda(n) = \max_{i=0,\ldots,n-1}(x_{i+1}-x_i)$. Then it is easy to show that $V\left(T_{i,i+1}(y,x_i,x_{i+1})\right) \leq \frac{\lambda(n)^2}{2}\sigma^2$ and the total variance is $V\left(T_{n+1}(y,a,b)\right) \leq \frac{\sigma^2}{2}\,n\lambda(n)^2 \underset{n\rightarrow\infty}{\longrightarrow} 0$ from our hypothesis. So the estimator is consistent. □

Lemma 2.3 We have proven in Lemma 2.2 that trapezoidal rule for the noisy data gives a consistent estimator for the trapezoidal estimation of the actual data, thus χ_l, χ_r are consistent estimators of the true x_l, x_r respectively, with relevant integrals trapezoidal calculated. If the interval $[a,b]$ is such that both $x_l, x_r \in [a,b]$ then *ESE* is a consistent estimator of trapezoidal calculated $x_S = \frac{x_l+x_r}{2}$. If $x_l > b$ then recalling Proof of Lemma 2.1 $s_l(a,x)$ is a decreasing function, so the minimum χ_l is achieved when $\chi_l = b$, the rightmost value of $[a,b]$. If $x_r < a$ then, recalling same Lemma, $s_r(b,x)$ is an increasing function, so the maximum χ_r is achieved when $\chi_r = a$, the leftmost value of $[a,b]$. Thus, for every possible case, *ESE* is a consistent estimator of the *TESE* given by integrals calculated via trapezoidal rule. □

Lemma 2.4 From Eq. 2 we see that $F^{'}(x) = f^{'}(x) - \frac{f(b)-f(a)}{b-a} = 0 \rightarrow f^{'}(x) = \frac{f(b)-f(a)}{b-a}$ while $F^{''}(x) = f^{''}(x)$, provided that smoothness conditions allows us to perform differentiations. Thus both necessary and sufficient conditions for extrema are satisfied. □

Lemma 2.5 For all $Y_j, j = 0, 1, \ldots, n$ it holds $E(Y_j) = F(x_j)$, so if we take the noisy data instead of accurate, we have $E\left(\underset{j\in[0,n]}{min}\{Y_j\}\right) = \underset{j\in[0,n]}{min}\{F(x_j)\}$ and $E\left(\underset{j\in[0,n]}{max}\{Y_j\}\right) = \underset{j\in[0,n]}{max}\{F(x_j)\}$ □

Lemma 2.6 Due to the zig–zag process we can accept that $\frac{y_i+y_{i-1}}{2}$ is approximately equal to both $f(x_i)$, $f(x_{i-1})$ and proceed to $y_i - y_{i-1} = y_i - \frac{y_i+y_{i-1}}{2} + \frac{y_i+y_{i-1}}{2} - y_{i-1} \approx (y_i - f(x_i)) - (y_{i-1} - f(x_{i-1})) = (-1)^i \epsilon_i + (-1)^{i+1} \epsilon_{i-1}$, so $\frac{y_i-y_{i-1}}{2} \approx \frac{1}{2}\left((-1)^i \epsilon_i + (-1)^{i+1} \epsilon_{i-1}\right)$ and $\left(\frac{y_i-y_{i-1}}{2}\right)^2 \approx \frac{1}{4}{\epsilon_i}^2 + \frac{1}{4}{\epsilon_{i-1}}^2 - \frac{1}{2}\epsilon_i\epsilon_{i-1}$. Now due to the *iid* property of our error terms it holds $\frac{2}{n}\sum_{i=1}^{n}\left(\frac{y_i-y_{i-1}}{2}\right)^2 \approx 2\left(\frac{1}{2}s^2\right) - 0 = s^2$ □

Lemma 2.7 From Def. 2.15, $\chi_D = \frac{1}{2}min(y_i - g(x_i)) + \frac{1}{2}max(y_i - g(x_i))$ and by using Eq. 28 we find $Var(\chi_D) = s_D^2 = \left(\frac{1}{2}\right)^2 Var(y_i) + \left(\frac{1}{2}\right)^2 Var(y_i) = \frac{1}{2}s^2$. □

REFERENCES

[1] W. Bardsley and R. Childs. Sigmoid curves, nonlinear double-reciprocal plots and allosterism. *Biochemical Journal*, 149:313–328, 1975.

[2] L. Bertalanffy. Principles and theory of growth. In *Fundamental Aspects of Normal and Malignant Growth*. New York: Elsevier, 1960.

[3] D. T. Christopoulos. *inflection: Finds the inflection point of a curve.*, 2013. Package version 1.1.

[4] D. T. Christopoulos. Extraction of the global absolute temperature for northern hemisphere using a set of 6190 meteorological stations from 1800 to 2013. *Journal of Atmospheric and Solar-Terrestrial Physics*, 128:70 – 83, 2015.

[5] I. C. Demetriou. L2CXCV: A fortran 77 package for least squares convex/concave data smoothing. *Computer Physics Communications*, 174(8):643 – 668, 2006.

[6] J. C. Fisher and R. H. Pry. A simple substitution model of technological change. *Technological Forecasting and Social Change*, 3:5–88, 1971.

[7] M. Friedman and L. Savage. The utility analysis of choices involving risk. *Journal of Political Economy*, 4:279–304, 1948.

[8] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical Transactions of the Royal Society of London*, 115:513–585, 1825.

[9] J. H. Han and T. Poston. Chord-to-point distance accumulation and planar curvature: a new approach to discrete curvature. *Pattern Recognition Letters*, 22:1133–1144, 2001.

[10] J. J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Scientists*, 81:3088–3092, 1984.

[11] S. Loncaric. A survey of shape analysis techniques. *Pattern Recognition*, 31:983–1001, 1998.

[12] C. Marchetti and N. Nakicenovic. The dynamics of energy systems and the logistic substitution model report. Technical report, International Institute for Applied Systems Analysis, RR-79-13, Laxenburg, Austria, 1979.

[13] J. Meddings, R. Scott, and G. Fick. Analysis and comparison of sigmoidal curves: application to dose-response data. *Am J Physiol Gastrointest Liver Physiol*, 257:982–989, 1989.

[14] T. Modis. *Predictions*. Simon and Schuster, New York, 1992.

[15] F. Mokhtarian and A. Mackworth. Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:34–43, 1986.

[16] M. Mokji and S. A. Bakar. Starfruit shape defect estimation based on concave and convex area of a closed planar curve. *Jurnal Teknologi*, 48:75–89, 2008.

[17] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2015.

[18] P. Turchin. *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. Princeton University Press, 2003.

[19] S. Upadhyay. *Chemical Kinetics and Reaction Dynamics*. Springer, 2006.

[20] P. Verhulst. Notice sur la loi que la population poursuit dans son accroissement. *Correspondance mathematique et physique*, 10:113–121, 1838.

[21] V. Volterra. *Lesons sur la theorie mathematique de la lutte pour la vie*. Gauthier-Villars Paris, 1931. Reissued 1990.

[22] M. Wand and M. Jones. *Kernel Smoothing*. Chapman and Hall, 1995.