*Adv. Appl. Prob.* **36,** 602–615 (2004)
*Printed in Northern Ireland*


# MODELLING THE PCR AMPLIFICATION PROCESS BY A SIZE-DEPENDENT BRANCHING PROCESS AND ESTIMATION OF THE EFFICIENCY

N. LALAM * AND
C. JACOB,** *Institut National de la Recherche Agronomique, Jouy-en-Josas*
P. JAGERS,**** *Chalmers University of Technology, Göteborg*

**Abstract**

We propose a stochastic modelling of the PCR amplification process by a size-dependent branching process starting as a supercritical Bienaymé–Galton–Watson transient phase and then having a saturation near-critical size-dependent phase. This model allows us to estimate the probability of replication of a DNA molecule at each cycle of a single PCR trajectory with a very good accuracy.

*Keywords:* Polymerase chain reaction; size-dependent branching; kinetic model; conditional least squares

2000 Mathematics Subject Classification: Primary 60J80; 62F12
Secondary 62P10

## 1. Introduction

The polymerase chain reaction (PCR; see [14]) is an *in vitro* enzymatic reaction allowing the amplification of the number of copies of a specific region of DNA. This technique is largely used in molecular biology [15] since it enables the detection of low abundance of DNA. The aim of quantitative PCR is to determine the initial amount of specific nucleic acids (the target) present in a sample. Quantitative analysis is now a major use of PCR and has many applications in virology [13], gene-expression studies [25] or pathogenic bacteria detection [9]. For more detailed applications, see [2].

The PCR proceeds through a succession of 30–50 cycles of replication. The number of copies of the target DNA increases at most by a factor of 2 at each cycle, but, in practice, the probability that a molecule will be duplicated after one cycle of amplification, known as the efficiency of the reaction, is less than 1. The early cycles of PCR are characterized by an exponential increase in the target population. Then, as the reaction components become limiting, the efficiency of the reaction decreases, leading to a saturation phase [21] decomposed into a linear phase and a phase called the plateau.

Our aim is to propose a modelling of the whole PCR amplification process in order to construct estimators of the efficiency from a single amplification trajectory. We choose to model the amplification process by a size-dependent branching process since branching process

Received 24 March 2003; revision received 9 December 2003.
* Current address: Eurandom, PO Box 513, 5600 MB Eindhoven, The Netherlands.
Email address: lalam@eurandom.tue.nl
** Postal address: INRA, Laboratoire de Biométrie, 78352 Jouy-en-Josas Cedex, France.
Email address: cj@banian.jouy.inra.fr
*** Postal address: Chalmers University of Technology, S-412 96 Göteborg, Sweden.
Email address: jagers@math.chalmers.se

theory [6] allows us to take into account the stochastic variability of the reaction and the size-dependent setting enables us to take into account the saturation phase of the amplification that appears much less noisy relative to the exponential one.

Branching processes have already been used to model the amplification process either for explaining the randomness of the amplification or for estimating some characteristics of this process. Single-type branching processes in discrete time were first applied to model the PCR for estimating replication errors of the Taq DNA polymerase [10]. Sun [24], Weiss and von Haeseler [26], Stolovitzky and Cecchi [23], Peccoud and Jacob [17] and Piau [19] have also used single-type branching process theory to model the exponential phase. Nedelman *et al.* [16] proposed a multitype Bienaymé–Galton–Watson branching process for native DNA, long products and short products. Relying on the enzymological approach of the PCR described by Schnell and Mendoza [22], Jagers and Klebaner [7] used a size-dependent branching process with the following efficiency for the reaction:

$$p(N_n) = \frac{K}{K + N_n}, \tag{1.1}$$

where $N_n$ is the number of DNA molecules present at cycle $n$ and $K$ is the Michaelis–Menten constant of the reaction. They obtained that $\lim_{n\to\infty} N_n/n = K$ almost surely, leading to a theoretical proof of the existence of the linear phase of the saturation observed in real PCR data.

Jacob and Peccoud [5] proved strong consistency and gave an exponential rate of convergence of the estimator of the efficiency from migrating binomial observations in the context of the modelling of the exponential phase by a supercritical Bienaymé–Galton–Watson process.

We describe in Section 2 the quantification method that is currently widely used.

In Section 3, we define a size-dependent modelling of the whole amplification process based on the concept of saturation: denoting the saturation threshold by $S$, we assume that the efficiency at cycle $n$ is a decreasing function of $\mathcal{S}(N_n)S^{-1}$ denoted by $f(\mathcal{S}(N_n)S^{-1})$, where $\mathcal{S}(N_n) = S\,\mathbf{1}_{\{N_n<S\}} + N_n\,\mathbf{1}_{\{N_n\geq S\}}$, $\mathbf{1}_{\{\cdot\}}$ being the indicator function. The threshold $S$ and the function $f(\cdot)$ are chosen in order to generalize (1.1): in (1.1), $S = N_0$, i.e. the saturation occurs at the beginning of the reaction, whereas in the model proposed here, there is an exponential phase which persists as long as $N_n \leq S$.

In Section 4, we study the asymptotic behavior of the size-dependent process relying on results of Kersting [8] and Jagers and Klebaner [7] and we prove that the efficiency model we propose is asymptotically valid in the linear part of the saturation phase. We will use this behavior to study the asymptotic properties of the estimator of $K$ given in Section 5 as $n \to \infty$ and based on the conditional least-squares method [11].

In Section 6, we study the properties of the estimators at finite distances relying on simulations and real PCR data: we estimate parameters of the efficiency model and also the last cycle of the exponential phase. Although the asymptotics concern the number of observed cycles and although we have only a few reliable observations, we obtain a very good accuracy for the estimations since the strong law of large numbers plays a significant role as soon as $n$ belongs to the end of the exponential phase or the beginning of the saturation phase: for example, relying on five successive observations of the beginning of the saturation phase, the standard deviation of the estimator of the efficiency of the exponential phase based on 500 noisy simulated trajectories is less than $10^{-3}$.

## 2. The usual quantitative PCR methodology

The most usual method for quantifying the initial amount of the target uses a standard which is co-amplified with the target [3]. Several dilutions $\{N_{0,i}\}_i$ of the standard with known initial number of DNA molecules are needed and the initial amount of the target is calculated from the determination of the efficiency $p$ in the exponential phase by using linear regression analysis based on a single observation in the exponential phase for each amplification trajectory. The quantification is based on the classical assumption that the fluorescence measured at cycle $n$, denoted by $F_n$, is proportional to the number of DNA molecules, $N_n$, the accumulated DNA molecules being measured thanks to the fluorescence they emit [4]. But this method has several drawbacks: it assumes that $p$ is identical for all the amplification trajectories and relies on the strong assumption that the target and the standard have identical efficiency, but some authors have questioned the validity of these hypotheses [18], [20]. Moreover, it uses only one observation in the exponential phase per amplification trajectory and the quantification by regression needs many observations (generally the 96 wells of the measuring apparatus). Furthermore, the quantification is based on the approximate relationship $N_n \simeq (1+p)^n N_0$, which leads to the confusion of the measurement error and the variability of the reaction in the regression model error.

## 3. Stochastic modelling of the amplification process

We assume that each molecule can give birth at the next cycle to two identical molecules if the replication succeeds or remains unchanged if the replication fails. The number of molecules at cycle $n+1$ is given by the recursion formula

$$N_{n+1} = \sum_{i=1}^{N_n} Y_{n+1,i}, \tag{3.1}$$

where $Y_{n+1,i}$ is the number of descendants at cycle $n+1$ of the $i$th molecule belonging to cycle $n$. We assume that the $\{Y_{n+1,i}\}_i$ are independent and identically distributed (i.i.d.) random variables conditionally to $\mathcal{F}_n$, the $\sigma$-algebra generated by $N_0, \dots, N_n$, and that the replication depends only on the initial conditions and on the amount of molecules already synthesized. Then the process may be considered as a size-dependent process and, denoting the offspring mean and variance by $m(N_n) = \mathrm{E}(Y_{n+1,i} \mid \mathcal{F}_n)$ and $\sigma^2(N_n) = \mathrm{var}(Y_{n+1,i} \mid \mathcal{F}_n)$ respectively, we have

$$\begin{gathered}
\mathrm{P}(Y_{n+1,i} = 2 \mid \mathcal{F}_n) = p(N_n), \\
\mathrm{P}(Y_{n+1,i} = 1 \mid \mathcal{F}_n) = 1 - p(N_n), \\
m(N_n) = 1 + p(N_n), \\
\sigma^2(N_n) = p(N_n)(1 - p(N_n)),
\end{gathered}$$

where $p(N_n)$ is the efficiency at cycle $n+1$.

We model here the entire amplification process by a single-type size-dependent branching process. The model is based on the following assumptions.

**Assumption 3.1.** *For all $n$, $\mathcal{K}_n + N_n = \mathcal{K}_0 + N_0$, where $\mathcal{K}_n$ is the experimental material as measured by the number of DNA molecules that could be synthesized.*

**Assumption 3.2.** *A saturation phenomenon occurs mainly because of a depletion of the primers.*

More precisely, there exists a saturation coefficient, denoted by $s$, such that, when $\mathcal{K}_n/N_n > s$, the amplification process is not limited, i.e. the underlying branching process may be considered as a supercritical Bienaymé–Galton–Watson process (exponential phase), whereas, as soon as $\mathcal{K}_n/N_n \leq s$, a saturation phenomenon occurs and the efficiency decreases. According to Assumption 3.1, let $S = (\mathcal{K}_0 + N_0)(s+1)^{-1}$; then

$$\frac{\mathcal{K}_n}{N_n} \leq s \iff \frac{N_n}{S} \geq 1. \tag{3.2}$$

If the saturation instead occurs because of a loss of activity of the DNA polymerase as $N_n$ increases, then we may assume that there exists a saturation threshold $S_0$ such that the saturation occurs as soon as $N_n - N_0 \geq S_0$, that is, $N_n \geq S$, where $S = N_0 + S_0$. Therefore (3.2) may be considered as a general assumption.

Let us recall that $\mathcal{S}(x) = S\,\mathbf{1}_{\{x<S\}} + x\,\mathbf{1}_{\{x\geq S\}}$. We assume that the efficiency $p(N_n)$ is a decreasing function of $\mathcal{S}(N_n)S^{-1}$, which we denote by $f(\mathcal{S}(N_n)S^{-1})$.

Let $n_{\mathrm{s}}$ be the last cycle of the nonsaturated phase: $n_{\mathrm{s}} = \sup\{n : N_{n-1} < S\}$. Then, since $\{N_n\}_n$ is increasing because $\mathrm{P}(Y_{n+1,i} = 0 \mid \mathcal{F}_n) = 0$, for all $n$,

$$f(\mathcal{S}(N_n)S^{-1}) = \begin{cases} f(1) & \text{if } N_n < S \text{ (equivalently, if } n \leq n_{\mathrm{s}} - 1\text{)},\\ f(N_n S^{-1}) & \text{if } N_n \geq S \text{ (equivalently, if } n \geq n_{\mathrm{s}}\text{)}. \end{cases}$$

Here and in the sequel, we will denote $f(1)$ by $p$. We will use the following model for $f(\mathcal{S}(N_n)S^{-1})$, generalizing (1.1):

$$f(\mathcal{S}(N_n)S^{-1}) = \frac{\mu}{\mu + \mathcal{S}(N_n)S^{-1}} \frac{1 + \exp(-C(\mathcal{S}(N_n)S^{-1} - 1))}{2}.$$

Let $K = \mu S$. Then $p(N_n)$ may be rewritten as

$$p(N_n) = \frac{K}{K + \mathcal{S}(N_n)} \frac{1 + \exp(-C(\mathcal{S}(N_n)S^{-1} - 1))}{2}. \tag{3.3}$$

The model (1.1) is a particular case of (3.3) with $S = N_0$ and $C = 0$.

**Remark 3.1.** When considering real PCR amplification trajectories with observations in fluorescence units and assuming that, for all $n$, the observed fluorescence $F_n$ is proportional to the associated number $N_n$ of DNA molecules, then the efficiency expressed in fluorescence units $p(F_n)$ may be obtained from the model (3.3) with $K$, $S$ and $\mathcal{S}(N_n)$ replaced by their equivalents in fluorescence units, $K_F$, $S_F$ and $\mathcal{S}_F(F_n) = S_F\,\mathbf{1}_{\{F_n<S_F\}} + F_n\,\mathbf{1}_{\{F_n\geq S_F\}}$. The parameters $K_F$ and $S_F$ are proportional to $K$ and $S$ and $\mathcal{S}(N_n)S^{-1} = \mathcal{S}_F(F_n)S_F^{-1}$.

**Remark 3.2.** Let $n_{\mathrm{s}}^{\mathrm{obs}}$ be the cycle of the end of the exponential phase, i.e. $n_{\mathrm{s}}^{\mathrm{obs}}$ is the last cycle $n$ such that $p(N_{n-1}) = p$. If $S < N_{n_{\mathrm{s}}}$, then $n_{\mathrm{s}}^{\mathrm{obs}} = n_{\mathrm{s}}$; if $S = N_{n_{\mathrm{s}}}$, then $n_{\mathrm{s}}^{\mathrm{obs}} = n_{\mathrm{s}} + 1$. The fact that $n_{\mathrm{s}}^{\mathrm{obs}}$ is $n_{\mathrm{s}}$ or $n_{\mathrm{s}} + 1$ was noticed by Peccoud and Jacob (private communication, 1998) from the observation of replicate PCR trajectories.

The offspring mean model $m(N) = 1 + p(N)$ may be rewritten as

$$m(N) = 1 + \frac{K_C}{\mathcal{S}(N)} + r(\mathcal{S}(N)), \tag{3.4}$$

where

$$K_C = K\left(\frac{1+\delta_C}{2}\right),$$
$$\delta_C = \mathbf{1}_{\{C=0\}},$$
$$r(\mathcal{S}(N)) = \frac{K}{\mathcal{S}(N)(K+\mathcal{S}(N))}\left[-K_C + \mathcal{S}(N)\frac{\exp(-C(\mathcal{S}(N)S^{-1}-1))-\delta_C}{2}\right].$$

Since $\lim_{x\to\infty} x(\exp(-C(x-1))-\delta_C) = 0$, it follows that $r(N) = O(N^{-2})$.

The offspring variance $\sigma^2(N)$ satisfies

$$\sigma^2(N) = \frac{K_C}{\mathcal{S}(N)} - r_+(\mathcal{S}(N)), \tag{3.5}$$

where

$$\begin{aligned} r_+(\mathcal{S}(N)) = {} & \frac{K}{\mathcal{S}(N)(K+\mathcal{S}(N))^2} \\ & \times \left[K\mathcal{S}(N)\left(\frac{\exp(-C(\mathcal{S}(N)S^{-1}-1))+1}{2}\right)^2\right. \\ & \left. + (K+\mathcal{S}(N))(K_C-\mathcal{S}(N))\frac{\exp(-C(\mathcal{S}(N)S^{-1}-1))-\delta_C}{2}\right]. \end{aligned} \tag{3.6}$$

## 4. Asymptotic behavior of the process

In the exponential phase, $m(N) = 1+p$ and $\sigma^2(N) = p(1-p)$. The Bienaymé–Galton–Watson phase is a transient phase of the model. Since $S$ is finite, the asymptotic behavior of the process is given by the near-critical saturation part of the model with offspring mean (3.4).

The model defined by (3.4) and (3.5) belongs to the class of models satisfying the following assumptions (these are modified versions of Assumptions 1.1–1.4 of [11]).

**Assumption 4.1.** *For all $N$,*

$$m(N) = 1 + p(N) \quad \text{and} \quad p(N) \le c(\mathcal{S}(N))^{-\alpha} \le cN^{-\alpha},$$

*where $\alpha = 1$ and $c < \infty$.*

**Assumption 4.2.** *There exists an $R$ such that, for all $N \ge R$,*

$$\sigma^2(N) \le K_C(\mathcal{S}(N))^{\beta} \le K_C N^{\beta},$$

*where $\beta = -1$.*

**Assumption 4.3.** *The function $N \mapsto p(N)$ decreases to $0$ and $N \mapsto \sigma^2(N)/N^2 p(N)$ is ultimately decreasing and satisfies*

$$\int_1^\infty \frac{\sigma^2(x)}{x^2 p(x)}\,\mathrm{d}x < \infty.$$

**Assumption 4.4.** *For $K > 0$ and $C \ge 0$,*

$$0 < p(N) < 1.$$

Let $\{a_n\}_n$ be the increasing sequence defined by $a_0 = 1$ and, for $n \geq 1$,

$$a_n = a_{n-1}(1 + f(\mathcal{S}(a_{n-1})S^{-1})). \tag{4.1}$$

Then $a_n = (1 + p)^n$ for $n \leq n_{\mathrm{s}}$.

Using Kersting's result [8], the process satisfies $\lim_{n\to\infty} N_n a_n^{-1} = 1$ almost surely.

Furthermore, as in Jagers and Klebaner's paper [7], (4.1) implies that

$$\frac{a_n}{n} = \frac{a_0}{n} + \frac{1}{n}\sum_{k=1}^{n} a_{k-1} f(\mathcal{S}(a_{k-1})S^{-1}). \tag{4.2}$$

According to Kersting, $\lim_{k\to\infty} a_k = \infty$, implying that $\lim_{k\to\infty} a_{k-1} f(\mathcal{S}(a_{k-1})S^{-1}) = K_C$. Then, thanks to the Toeplitz lemma, (4.2) implies that $\lim_{n\to\infty} a_n n^{-1} = K_C$. Consequently, $\lim_{n\to\infty} N_n n^{-1} = K_C$ almost surely, entailing that $N_n = K_C n + o(n)$, which is consistent with the linear part of the saturation phase. Hence, this model asymptotically fits the linear part of the saturation phase.

## 5. Estimation of the efficiency

We estimate the unknown parameters $K_C$, $S$ and $C$ in (3.4) and therefore the efficiency $p = K(K + S)^{-1} = \mu(\mu + 1)^{-1}$ of the exponential phase from a single PCR trajectory by using the nonlinear autoregressive model deduced from (3.1):

$$N_k = m(N_{k-1})N_{k-1} + \eta_k, \qquad m(N_{k-1}) = 1 + f(\mathcal{S}(N_{k-1})S^{-1}), \tag{5.1}$$

where the mean and variance of $\eta_k$ conditionally on $N_{k-1}$ are

$$\mathrm{E}(\eta_k \mid N_{k-1}) = 0, \qquad \sigma^2(\eta_k \mid N_{k-1}) = N_{k-1}\sigma^2(N_{k-1}). \tag{5.2}$$

According to (3.5),

$$\sigma^2(\eta_k \mid N_{k-1}) = N_{k-1}\left(\frac{K_C}{\mathcal{S}(N_{k-1})} - r_+(\mathcal{S}(N_{k-1}))\right).$$

In order to define the optimal estimator not only from an asymptotic point of view but also from a finite-distance point of view, we generalize the optimal contrast with weights $\{N_{k-1}^{-1} N_{k-1}^{-\beta}\}_k$, which were defined in [11] for studying only the asymptotic properties, by taking into account the dependence of $\sigma^2(\eta_k \mid N)$ on $\mathcal{S}(N)$. The estimator of $(K_C, S, C)$ will therefore minimize the following conditional least-squares equation:

$$SS_{h,n}(K_C, S, C) = \sum_{k=h+1}^{n} (N_k - (1 + f(\mathcal{S}(N_{k-1})S^{-1}))N_{k-1})^2 N_{k-1}^{-1} \mathcal{S}(N_{k-1})^{-\beta}, \tag{5.3}$$

where $\beta = -1$. Let

$$((\widehat{K_C})_{h,n}, \hat{S}_{h,n}, \hat{C}_{h,n}) = \arg\min_{K_C,S,C} SS_{h,n}(K_C, S, C).$$

Since $S$ is unknown, we replace the weights $\mathcal{S}(N_{k-1})^{-\beta}$ of the contrast (5.3) by the weights $\hat{\mathcal{S}}(N_{k-1})^{-\beta}$, where $S$ is estimated by $N_{n_s}$. Therefore, (5.3) becomes

$$SS_{h,n}(K_C, S, C) = \sum_{k=h+1}^{n_s} (N_k - (1 + f(1))N_{k-1})^2 N_{k-1}^{-1} N_{n_s} + \sum_{k=n_s+1}^{n} (N_k - (1 + f(N_{k-1}S^{-1}))N_{k-1})^2 \tag{5.4}$$

and (5.4) is asymptotically equivalent to (5.3). The amplification-rate model (3.4) satisfies Assumption 1.1′ of [11]: $m(N) = 1 + K_C(\mathcal{S}(N))^{-1} + O(\mathcal{S}(N)^{-\bar{\alpha}})$, where $\bar{\alpha} = 2$. But according to (3.4) and the results of [11] obtained in the general context of size-dependent branching processes, the consistency of $\{((\widehat{K_C})_{h,n}, \hat{S}_{h,n}, \hat{C}_{h,n})\}_n$ needs at least the asymptotic identifiability of $(K_C, S, C)$ in $m(\cdot)$ at the rate $N_{k-1}^{\bar{\alpha}}$ and requires that $\beta + 2\bar{\alpha} \leq 1$, which is not the case since $\beta = -1$ and $\bar{\alpha} = 2$. According to (3.4), $K_C$ is identifiable at the rate $N_{k-1}^{\alpha}$ with $\alpha = 1$ and $(S, C)$ appears only in the negligible functions $r(\cdot)$ and $r_+(\cdot)$. Therefore, we will study the properties of $\{(\widehat{K_C})_{h,n}\}_n$ and we will consider $\{(\hat{S}_{h,n}, \hat{C}_{h,n})\}_n$ as a nuisance parameter, which we denote by $\nu$. We write $SS_{h,n,\nu}(\theta)$ instead of $SS_{h,n}(K_C, S, C)$, where $\theta = K_C$; we are studying the properties of the estimator of the parameter $\theta$. We have

$$\hat{\theta}_{h,n,\nu} = \arg\min_{\theta} SS_{h,n,\nu}(\theta). \tag{5.5}$$

Our asymptotic study concerns $n \to \infty$. Notice that, in the PCR setting, $n \to \infty$ entails that $N_n \to \infty$ almost surely.

We assume that $\theta = K_C$ belongs to a compact set $\Theta$ in $\mathbb{R}^+ \setminus \{0\}$ and $\theta_0 \in \overset{\circ}{\Theta}$. Set $\hat{\mathcal{S}}(a_{k-1}) = a_{n_s} = (1 + p)^{n_s}$ for all $k \leq n_s$ and $\hat{\mathcal{S}}(a_{k-1}) = a_{k-1}$ for all $k \geq n_s + 1$.

Let

$$\Phi_{h,n}^{-1}(n_s) = \sqrt{\sum_{k=h+1}^{n} a_{k-1}\hat{\mathcal{S}}(a_{k-1})^{-\beta-2\alpha}} = \sqrt{\sum_{k=h+1}^{n_s} (1 + p)^{k-1-n_s} + n - n_s}.$$

**Proposition 5.1.** *Let $h$ be fixed. Then* $\lim_{n\to\infty} \hat{\theta}_{h,n,\nu} = K_C$ *almost surely and*

$$\lim_{n\to\infty} \Phi_{h,n}^{-1}(n_s)(\hat{\theta}_{h,n,\nu} - K_C) \overset{\mathrm{D}}{=} \mathrm{N}(0, K_C). \tag{5.6}$$

**Remark 5.1.** As a direct consequence, when $S$ is known, this result allows us to construct a confidence interval for $p = \mu(\mu + 1)^{-1}$ or a test of equality of the efficiencies of two trajectories since (5.6) implies that

$$\lim_{n\to\infty} \Phi_{h,n}^{-1}(n_s)((\widehat{\mu_C})_{h,n,\nu} - \mu_C) \overset{\mathrm{D}}{=} \mathrm{N}(0, \mu_C S^{-1}), \tag{5.7}$$

where $\mu_C = K_C S^{-1}$. According to (5.7), the larger is $S$, the better is the accuracy of $\hat{p}_{h,n,\nu}$. For example, for the amplification well 21 of data set 1 (see Section 6), if $S_F$ is assumed known ($S_F = (\widehat{S_F})_{h,n}$), we obtain the following approximate 95% confidence interval for $p$: [0.8435983, 0.8435994].

*Proof.* First, notice that, since $S$ is fixed, $n_{\mathrm{s}}$ is almost surely finite since $\lim_{n\to\infty} N_n = \infty$ almost surely. Second, notice that, since the contrast $SS_{h,n,\nu}(\theta)$ is a generalization of the contrast studied in [11], in Conditions 3.7, 3.8, 4.1–4.3 of [11], we can replace the quantities $a_{k-1}^{\beta}$, $a_{k-1}^{\alpha_*}$ and $a_{k-1}^{\alpha_{**}}$ by the quantities $\hat{\mathcal{S}}(a_{k-1})^{\beta}$, $\hat{\mathcal{S}}(a_{k-1})^{\alpha_*}$ and $\hat{\mathcal{S}}(a_{k-1})^{\alpha_{**}}$ in order to prove the consistency and the rate of convergence of the estimators, where $\alpha_* = \alpha$ and $\alpha_{**} = \bar{\alpha}$.

To get the strong consistency, we apply Corollary 3.1 of [11] since the contrast $SS_{h,n,\nu}(\theta)$ is asymptotically equivalent on each trajectory to the contrast where the weights are $N_{k-1}^{-1-\beta}$. The model (3.4) satisfies Conditions 3.6 and 3.7 of [11] for $\gamma = 1+\beta$ and $h$ fixed, leading to the strong consistency of $\{\hat{\theta}_{h,n,\nu}\}_n$, since

$$\sup_{\nu}\sup_{\theta} |r_{\theta,\nu}(N)|N^2 < \infty, \quad \text{where } r_{\theta_0,\nu}(N) = r(N) \text{ (proving Condition 3.6 of [11])},$$

$$\lim_{n\to\infty} \sum_{k=h+1}^{n_{\mathrm{s}}} (1+p)^{k-1-n_{\mathrm{s}}} + n - n_{\mathrm{s}} = \infty \quad \text{(proving Condition 3.7(i) of [11])},$$

$$\sum_{k=n_{\mathrm{s}}+1}^{\infty} \bigg[\sum_{l=h+1}^{n_{\mathrm{s}}} (1+p)^{l-1-n_{\mathrm{s}}} + k - n_{\mathrm{s}}\bigg]^{-2} < \infty \quad \text{(proving Condition 3.7(ii) of [11])}.$$

Let us now prove (5.6). In view of (3.5), $\sigma^2(N) = K_C N^{\beta} - r_+(N)$ for $N$ large enough, where $\beta = -1$ and $r_+(N)$ defined in (3.6) is such that $0 \le r_+(N) = O(N^{\bar{\beta}})$ with $\bar{\beta} = -2$; this is Assumption 1.2′ of [11].

Let us check the Conditions 3.8 and 4.1–4.3 of [11] to get the rate of convergence on the set $\{n_{\mathrm{s}} = l\}$:

1. Condition 3.8. For all $N$,

$$\sup_{\theta,\nu} |r'_{\theta,\nu}(N)|N^2 < \infty \quad \text{and} \quad \sup_{\theta,\nu} |r''_{\theta,\nu}(N)|N^2 < \infty,$$

   where ′ denotes the derivative with respect to $\theta = K_C$.

2. Condition 4.1. Letting

$$B_n = \sqrt{\sum_{k=h+1}^{n} a_{k-1}\hat{\mathcal{S}}(a_{k-1})^{-\beta-2\alpha}},$$

   we have

$$\lim_{n\to\infty} B_n = \lim_{n\to\infty} \sqrt{\sum_{k=h+1}^{n_{\mathrm{s}}} (1+p)^{k-1-n_{\mathrm{s}}} + n - n_{\mathrm{s}}} = \infty.$$

3. Condition 4.2. We have

$$\lim_{n\to\infty} \bigg[\sum_{k=n_{\mathrm{s}}+1}^{n} a_{k-1}^{-1}\bigg]\bigg[\sum_{k=h+1}^{n_{\mathrm{s}}} (1+p)^{k-1-n_{\mathrm{s}}} + n - n_{\mathrm{s}}\bigg]^{-1/2} = 0.$$

   Since $a_n = K_C n + o(n)$,

$$\frac{\sum_{k=n_{\mathrm{s}}+1}^{n} a_{k-1}^{-1}}{\sqrt{\sum_{k=h+1}^{n_{\mathrm{s}}} (1+p)^{k-1-n_{\mathrm{s}}} + n - n_{\mathrm{s}}}} \le \sup_{k}\bigg|\frac{1}{K_C + o(k)/k}\bigg| \frac{\sum_{k=n_{\mathrm{s}}}^{n-1} k^{-1}}{\sqrt{n-n_{\mathrm{s}}}}.$$

Now

$$\sum_{k=n_{\mathrm{s}}}^{n-1} \frac{1}{k} \leq \int_{n_{\mathrm{s}}-1}^{n-1} \frac{\mathrm{d}t}{t} \leq \log(n-1) - \log(n_{\mathrm{s}}-1).$$

Thus, for fixed $h$,

$$\lim_{n\to\infty} \frac{\sum_{k=n_{\mathrm{s}}}^{n-1} k^{-1}}{\sqrt{n-n_{\mathrm{s}}}} = 0,$$

as required.

4. Condition 4.3. For all $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \sup_{n_{\mathrm{s}}+1\leq k\leq n} \mathrm{E}(R_k^2 \mathbf{1}_{\{R_k^2 \geq B_{k,n}^2 x^2\}} \mid \mathcal{F}_{k-1}) = 0$$

almost surely, where $R_k = \bar{Y}_k p'(N_{k-1}) N_{k-1} a_{k-1}^{1/2}$, $\bar{Y}_k = Y_{k,1} - 1 - p(N_{k-1})$ and $B_{k,n} = B_n a_{k-1}^{1/2}$. Let $g_\nu(x) = (\exp(-C(xS^{-1}-1)) + 1)/2$; then

$$p'(N_{k-1}) = \frac{2N_{k-1}(1+\delta_C)}{(2K_C + N_{k-1}(1+\delta_C))^2} g_\nu(N_{k-1}).$$

Since $p'(N_{k-1})N_{k-1} \leq 2$ and $\bar{Y}_k^2 \leq 1$, we have $R_k^2 \leq 4a_{k-1}$, yielding that, for all $x \in \mathbb{R}$,

$$\sup_{n_{\mathrm{s}}+1\leq k\leq n} \mathrm{E}(R_k^2 \mathbf{1}_{\{R_k^2 \geq B_{k,n}^2 x^2\}} \mid \mathcal{F}_{k-1}) \leq 4a_{n-1} \mathrm{P}(4 \geq B_n^2 x^2).$$

On the set $\{n_{\mathrm{s}} = l\}$, $\mathrm{P}(4 \geq B_n^2 x^2) = 0$ for all $n$ large enough since $\lim_{n\to\infty} B_n = \infty$, implying that

$$\lim_{n\to\infty} \sup_{n_{\mathrm{s}}+1\leq k\leq n} \mathrm{E}(R_k^2 \mathbf{1}_{\{R_k^2 \geq B_{k,n}^2 x^2\}} \mid \mathcal{F}_{k-1}) = 0$$

almost surely, as required.

Then, according to Proposition 4.1 of [11], on the set $\{n_{\mathrm{s}} = l\}$, (5.6) holds.

In view of Corollary 3.4.34 of [1], $\lim_{n\to\infty} \Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \stackrel{\mathrm{D}}{=} \mathrm{N}(0,\theta_0)$ is equivalent to $\lim_{n\to\infty} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x) = \mathrm{P}(X \leq x)$ for all real $x$, where $X \stackrel{\mathrm{D}}{=} \mathrm{N}(0,\theta_0)$. Then

$$\begin{aligned}
&\left|\mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x) - \sum_{l=0}^{N} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x \mid n_{\mathrm{s}} = l)\,\mathrm{P}(n_{\mathrm{s}} = l)\right| \\
&\quad \leq \sum_{l=N+1}^{\infty} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x \mid n_{\mathrm{s}} = l)\,\mathrm{P}(n_{\mathrm{s}} = l) \\
&\quad \leq \mathrm{P}(n_{\mathrm{s}} > N).
\end{aligned}$$

But

$$\begin{aligned}
\lim_{n\to\infty} &\sum_{l\leq N} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x \mid n_{\mathrm{s}} = l)\,\mathrm{P}(n_{\mathrm{s}} = l) \\
&= \sum_{l\leq N} \lim_{n\to\infty} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu} - \theta_0) \leq x \mid n_{\mathrm{s}} = l)\,\mathrm{P}(n_{\mathrm{s}} = l) \\
&= \mathrm{P}(X \leq x)\,\mathrm{P}(n_{\mathrm{s}} \leq N).
\end{aligned}$$

This yields that

$$\left|\lim_{n\to\infty} \mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu}-\theta_0)\le x) - \mathrm{P}(X\le x)\,\mathrm{P}(n_{\mathrm{s}}\le N)\right| \le \mathrm{P}(n_{\mathrm{s}}>N).$$

The right-hand side of the last inequality tends to 0 as $N\to\infty$ since $n_{\mathrm{s}}$ is almost surely finite on the nonextinction set. As a consequence, $\lim_{n\to\infty}\mathrm{P}(\Phi_{h,n}^{-1}(n_{\mathrm{s}})(\hat{\theta}_{h,n,\nu}-\theta_0)\le x) = \mathrm{P}(X\le x)\lim_{N\to\infty}\mathrm{P}(n_{\mathrm{s}}\le N) = \mathrm{P}(X\le x)$ and (5.6) is valid.

## 6. Study of the efficiency estimator at finite distances

We study the behavior of the estimator of the efficiency at finite distances using simulations and real PCR data.

The random variable $\sum_{i=1}^{N_n}(Y_{n+1,i}-1)$ has a binomial distribution, with parameters $N_n$ and $p(N_n)$, denoted by $\mathrm{bin}(N_n,\,p(N_n))$. This entails that $N_{n+1}$ is given by the recursion formula

$$N_{n+1} = N_n + \mathrm{bin}(N_n,\,p(N_n)).$$

To reproduce a sequence of real observations $\{F_n/F_{n-1}\}_n$, we add noise to the size-dependent branching process: $X_k = N_k + \varepsilon_k$, where $\{\varepsilon_k\}_k$ is assumed to be independent of $\{N_k\}_k$ and is a sequence of i.i.d. centered Gaussian random variables with variance $\sigma_\varepsilon^2$. The simulations were implemented in MATHEMATICA® (Wolfram Research Inc.).

### 6.1. Study of the bias and the variance of the estimator

Simulations in the context of the efficiency model (3.3) were performed with $N_0 = 1000$, $K = 4\times 10^{10}$, $S = 10^{10}$, $C = 0$ and either $\sigma_\varepsilon = 0$ or $\sigma_\varepsilon = 4\times 10^7$. The true efficiency of the exponential phase is $p = 0.8$. For each of these sets of parameter values, 500 trajectories of the process were simulated up to cycle 40. The series of $\hat{p}_{h,n,\nu} = \hat{K}_{h,n,\nu}/(\hat{K}_{h,n,\nu}+\hat{S}_{h,n})$ were computed for each trajectory and the empirical mean and variance of the estimator were calculated over the 500 trajectories all satisfying $n_{\mathrm{s}} = 28$ for different $h$ and $n$ (see Tables 1 and 2). We chose $h$ large enough for $X_h$ to be above the background noise.

Considering the non-noisy set of simulations, the bias equals zero and the variance is of maximum order $10^{-11}$. With the noisy simulations, the bias is smaller than $10^{-3}$ and the variance is a bit greater than in the non-noisy simulations. The comparison of the quantities of Tables 1 and 2 for sets of observations in different phases and such that $n-h=4$ shows that the estimator is a bit better in the saturation phase, whereas the increasing of $n-h$ due

TABLE 1: Mean and variance when $N_0 = 1000$, $p = 0.8$, $C = 0$, $\sigma_\varepsilon = 0$.

| $h$ | $n$ | $\mathrm{E}(\hat{p}_{h,n,\nu})$ | $\mathrm{var}(\hat{p}_{h,n,\nu})$ | $h$ | $n$ | $\mathrm{E}(\hat{p}_{h,n,\nu})$ | $\mathrm{var}(\hat{p}_{h,n,\nu})$ |
|---|---|---|---|---|---|---|---|
| 24 | 28 | 0.8 | $1.08654\times 10^{-11}$ | 24 | 28 | 0.8 | $1.08654\times 10^{-11}$ |
| 26 | 30 | 0.8 | $2.61273\times 10^{-12}$ | 24 | 30 | 0.8 | $2.46202\times 10^{-12}$ |
| 28 | 32 | 0.8 | $8.80084\times 10^{-13}$ | 24 | 32 | 0.8 | $8.53060\times 10^{-13}$ |

TABLE 2: Mean and variance when $N_0 = 1000$, $p = 0.8$, $C = 0$, $\sigma_\varepsilon = 4\times 10^7$.

| $h$ | $n$ | $\mathrm{E}(\hat{p}_{h,n,\nu})$ | $\mathrm{var}(\hat{p}_{h,n,\nu})$ | $h$ | $n$ | $\mathrm{E}(\hat{p}_{h,n,\nu})$ | $\mathrm{var}(\hat{p}_{h,n,\nu})$ |
|---|---|---|---|---|---|---|---|
| 24 | 28 | 0.799818 | $3.80016\times 10^{-5}$ | 24 | 28 | 0.799818 | $3.80016\times 10^{-5}$ |
| 26 | 30 | 0.800019 | $1.78158\times 10^{-6}$ | 24 | 30 | 0.799958 | $1.90581\times 10^{-6}$ |
| 28 | 32 | 0.799991 | $1.18139\times 10^{-7}$ | 24 | 32 | 0.799988 | $1.23439\times 10^{-7}$ |

to the use of the saturation phase allows us to increase the accuracy of the estimator. The fact that the bias and variance are very small and decrease as $n$ increases, with $n-h$ remaining fixed and small, may be explained by the large amount of molecules replicated in the cycles we have considered so that the strong law of large numbers due to the almost-sure convergence of $N_n a_n^{-1}$ with $\lim_{n\to\infty} a_n = \infty$ and the almost-sure convergence to 0 of $\varepsilon_n a_n^{-1}$ play a significant role at each cycle. Furthermore, relying only on the first five observations of the saturation phase of the noisy simulations ($(h,n)=(28,32)$), the standard deviation of $\hat{p}_{h,n,\nu}$ is less than $10^{-3}$: we get an accurate estimation of $p$ based only on the beginning of the saturation phase.

### 6.2. Study of the estimator accuracy on single trajectories

We study the efficiency estimator on single trajectories based on the estimation model (3.4). These trajectories will be either simulation trajectories described in Subsection 6.1 or real PCR amplification trajectories. While considering real PCR data, we will study two data sets; data set 1 was described in [18] and data set 2 was provided by the Laboratory of Phytopathology and Methodology of Detection, INRA.

Let

$$\overline{SS}_{h,n}(K_C, S, C) = \frac{SS_{h,n}(K_C, S, C)}{n-h},$$

where $SS_{h,n}(K_C, S, C)$ is defined by (5.4). From a theoretical point of view, we should normalize the contrast with $\sum_{k=h+1}^{n_s}(1+p)^{k-1-n_s} + n - n_s$, but, since this quantity depends on unknown parameters, we use the normalization $n-h$. Let $[h_0, n_0]$ be a large window of reliable observations. We search for the largest window $[h,n] \subset [h_0, n_0]$ such that the model is adequate, i.e. $\overline{SS}_{h,n}((\widehat{K_C})_{h,n}, \hat{S}_{h,n}, \hat{C}_{h,n})$ is minimum among the set

$$\{\overline{SS}_{h',n'}((\widehat{K_C})_{h',n'}, \hat{S}_{h',n'}, \hat{C}_{h',n'})\}_{h_0 \le h' < \hat{n}_s^{\text{obs,graph}} < n' \le n_0}, \tag{6.1}$$

where $\hat{n}_s^{\text{obs,graph}}$ is a graphical estimation of the end of the exponential phase defined as the first cycle of the decrease of ten consecutive values of the simple estimator of the amplification rate $\{X_k/X_{k-1}\}_k$ [17]. We set $h_0 = \sup\{k : X_{k-1} < 0\}$ since the values of the measurements of the emitted fluorescence have a meaning only when they are positive. Taking $[h,n]$ to realize the minimum of (6.1), we will estimate $n_s = \sup\{k : X_{k-1} < S\}$ by $\hat{n}_s = \sup\{k : X_{k-1} < \hat{S}_{h,n}\}$.

6.2.1. *Simulation and estimation using the model (3.4) with $C=0$.* Let $(h_0, n_0) = (26, 35)$. Then $(h,n) = (h_0, n_0) = (26, 35)$, which is consistent with the increase of the accuracy estimator as the number of observations increases when the estimation model is valid on all of the trajectory. We have $\hat{K}_{h,n,\nu} = 4.00311 \times 10^{10}$ and $\hat{S}_{h,n} = 10^{10}$. In Figure 1, the dotted line is the plot of $X_k/X_{k-1} - 1$ versus $k$ and the solid line is the plot of the estimation of $p(X_{k-1})$ versus $k$. We obtain a very accurate estimation of the efficiency.

6.2.2. *Real PCR data with the estimation model (3.4).* We study the two experimental data sets. We recall that $F_k$, which is the measured fluorescence at cycle $k$, is assumed proportional to $N_k$. For $h$ and $n$ realizing the minimum of (6.1), the dotted line in Figures 2 and 3 plots the estimation of the observed efficiency $F_k/F_{k-1} - 1$ versus $k$ and the solid line plots the estimated efficiency $\hat{p}(F_{k-1})$ versus $k$, where $\hat{p}(F_{k-1})$ is the estimation of the efficiency model $p(F_{k-1})$. We get an accurate estimation of the efficiency from the end of the exponential phase until the linear phase of the saturation, and then there is an over-estimation in the plateau phase.

Figure 2 gives the results obtained for well 21 of data set 1 ($(h_0, n_0) = (14, 29)$) and Figure 3 gives the results for well 16 of data set 2 ($(h_0, n_0) = (17, 30)$). Denote $\hat{K}_{h,n,\nu}$ and $\hat{S}_{h,n}$ when expressed in fluorescence units by $(\widehat{K_F})_{h,n,\nu}$ and $(\widehat{S_F})_{h,n}$ respectively.

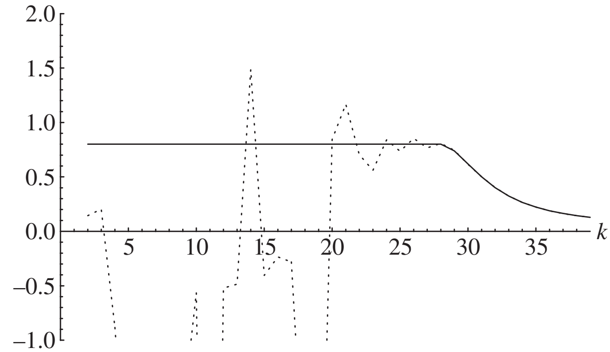


FIGURE 1: Simulation, with $h = 26$, $n = 35$, $\hat{p}_{h,n,\nu} = 0.800125$.

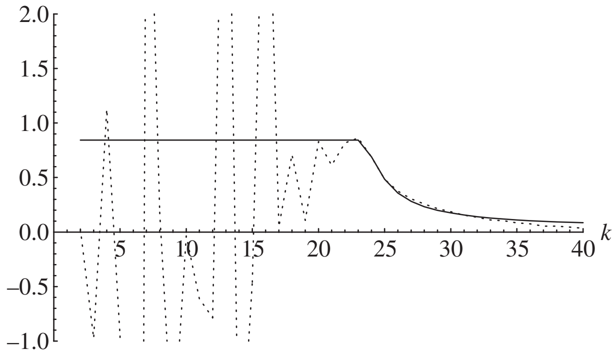


FIGURE 2: Well 21 of data set 1, with $h = 21$, $n = 25$, $\hat{n}_{\mathrm{s}} = 23$, $(\widehat{K_F})_{h,n,\nu} = 0.38055$, $(\widehat{S_F})_{h,n} = 0.070553$, $\hat{C}_{h,n} = 0.6$, $\hat{p}_{h,n,\nu} = 0.843599$.

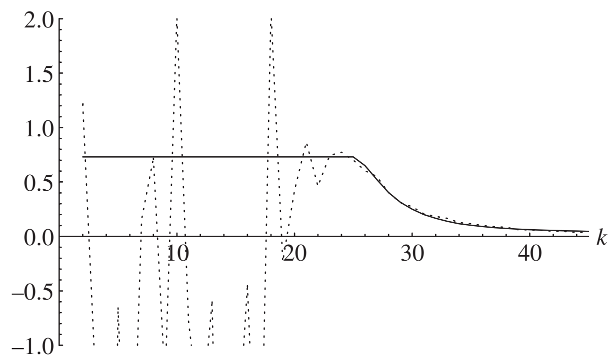


FIGURE 3: Well 16 of data set 2, with $h = 22$, $n = 29$, $\hat{n}_{\mathrm{s}} = 25$, $(\widehat{K_F})_{h,n,\nu} = 0.254064$, $(\widehat{S_F})_{h,n} = 0.0935645$, $\hat{C}_{h,n} = 0.07$, $\hat{p}_{h,n,\nu} = 0.730849$.

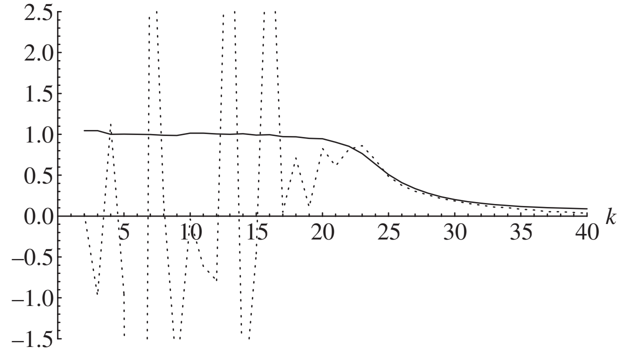


FIGURE 4: Well 21 of data set 1, with $h = 23$, $n = 27$, $(\widehat{K_F})_{h,n,\nu} = 0.22769$.

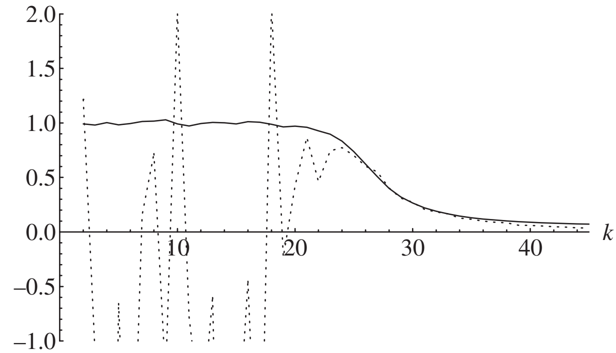


FIGURE 5: Well 16 of data set 2, with $h = 22$, $n = 29$, $(\widehat{K_F})_{h,n,\nu} = 0.216718$.

6.2.3. *Real PCR data with the estimation model (1.1).* In the efficiency model (1.1), the only unknown parameter is $K$. We observe a systematic bias which consists in an under-estimation of the rate of decrease of the efficiency at the beginning of the saturation phase of the model (3.4). Furthermore, as in Section 6.2.2, there is an over-estimation in the last cycles of the saturation phase (plateau phase). But the fit is relatively good for the other cycles, which validates the efficiency model studied by Jagers and Klebaner [7].

Figure 4 gives the result for well 21 of data set 1 and Figure 5 that for well 16 of data set 2. Other estimations relying on simulations and real PCR data can be found in [12].

## 7. Conclusion

To conclude, we give some prospectives about possible use of this work. It would be interesting to test the equality between efficiency of different amplification trajectories using the law of $(\widehat{\mu_C})_{h,n,\nu}$ given in Section 5 in order to validate or invalidate the basic assumption of equality of all the efficiencies for trajectories realized at the same time with the same measuring apparatus. More accurate results would need the knowledge of the distribution of the measurement error. The confidence interval of $p$ that is less than $10^{-5}$ (see Remark 5.1), calculated under the assumptions of null measurement error and knowledge of $S_F$, is consistent with the observed variance of $\hat{p}_{h,n,\nu}$ (Table 1). Therefore, according to Table 2, we may expect for real PCR data a confidence interval of maximum order $10^{-3}$.

Relying on the modelling of the exponential phase by a supercritical Bienaymé–Galton–Watson process, Jacob and Peccoud [5] built an asymptotic confidence interval of $N_0$, as $n \to \infty$. Since, for real-time PCR data, the observed fluorescence is less noisy in the saturation phase, it would be interesting to extend the results obtained in [5] to an estimator $\hat{F}_{0,n}$ of $F_0$ with $n$ belonging to the linear phase. When a calibration function making the conversion of fluorescent units into DNA molecules number is available, the confidence interval for $F_0$ would entail a confidence interval for $N_0$. Notice that relative quantitative PCR between two populations is now possible thanks to the relationship $\hat{F}^{(1)}_{0,n_1}/\hat{F}^{(2)}_{0,n_2} = \hat{N}^{(1)}_{0,n_1}/\hat{N}^{(2)}_{0,n_2}$.

## References


[1] Dacunha-Castelle, D. and Duflo, M. (1982). *Probabilités et Statistiques*, Tome 1. Masson, Paris.

[2] Ferré, F. (ed.) (1998). *Gene Quantification.* Birkhäuser, Boston.

[3] Gilliland, G., Perrin, S., Blanchard, K. and Bunn, H. F. (1990). Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *Proc. Nat. Acad. Sci. USA* **87,** 2725–2729.

[4] Higuchi, R., Dollinger, G., Walsh, P. S. and Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* **10,** 413–417.

[5] Jacob, C. and Peccoud, J. (1998). Estimation of the parameters of a branching process from migrating binomial observations. *Adv. Appl. Prob.* **30,** 948–967.

[6] Jagers, P. (1975). *Branching Processes with Biological Applications.* John Wiley, London.

[7] Jagers, P. and Klebaner, F. C. (2003). Random variation and concentration effects in PCR. *J. Theoret. Biol.* **224,** 299–304.

[8] Kersting, G. (1990). Some properties of stochastic difference equations. In *Stochastic Modelling in Biology*, ed. P. Tautu, World Scientific, Singapore, pp. 328–339.

[9] Kimura, B., Kawasaki, S., Nakano, H. and Fujii, T. (2001). Rapid, quantitative PCR monitoring of growth of clostridium botulinum type E in modified-atmosphere-packaged fish. *Appl. Environ. Microbiol.* **67,** 206–216.

[10] Krawczak, M., Reiss, J., Schmidtke, J. and Rosler, U. (1989). Polymerase chain reaction: replication errors and reliability of gene diagnosis. *Nucleic Acids Res.* **17,** 2197–2201.

[11] Lalam, N. and Jacob, C. (2003). Estimation of the offspring mean in a supercritical or near-critical size-dependent branching process. *Adv. Appl. Prob.* **36,** 582–601.

[12] Lalam, N. and Jacob, C. (2003). Modelling the PCR amplification process with size-dependent branching processes and estimation of the efficiency. Tech. Rep., Applied Mathematics and Informatics, INRA, Jouy-en-Josas.

[13] Mackay, I. M., Arden, K. E. and Nitsche, A. (2002). Real-time PCR in virology. *Nucleic Acids Res.* **30,** 1292–1305.

[14] Mullis, K. B. and Faloona, F. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods Enzymol.* **155,** 335–350.

[15] Mullis, K. B., Ferré, F. and Gibbs, R. A. (1994). *The Polymerase Chain Reaction.* Birkhäuser, Boston.

[16] Nedelman, J., Heagerty, P. and Lawrence, C. (1992). Quantitative PCR: procedures and precision. *Bull. Math. Biol.* **54,** 477–502.

[17] Peccoud, J. and Jacob, C. (1996). Theoretical uncertainty of measurements using quantitative polymerase chain reaction. *Biophys. J.* **71,** 101–108.

[18] Peccoud, J. and Jacob, C. (1998). Statistical estimations of PCR amplification rates. In *Gene Quantification*, ed. F. Ferré, Birkhäuser, Boston.

[19] Piau, D. (2001). Processus de branchement et champ moyen. *Adv. Appl. Prob.* **33,** 391–403.

[20] Raeymakers, L. (1995). A commentary on the practical applications of competitive PCR. *Genome Res.* **5,** 91–94.

[21] Saiki, R. K. *et al.* (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–491.

[22] Schnell, S. and Mendoza, C. (1997). Enzymological considerations for a theoretical description of the quantitative competitive polymerase chain reaction. *J. Theoret. Biol.* **184,** 433–440.

[23] Stolovitzky, G. and Cecchi, G. (1996). Efficiency of DNA replication in the polymerase chain reaction. *Biophysics* **93,** 12947–12952.

[24] Sun, G. (1995). The PCR and branching processes. *J. Comput. Biol.* **2,** 63–86.

[25] Vandenbroucke, I. I., Vandesempele, J., De Paepe, A. and Messiaen, L. (2001). Quantification of splice variants using real-time PCR. *Nucleic Acids Res.* **29,** e68.

[26] Weiss, G. and Von Haeseler, A. (1995). Modeling the PCR. *J. Comput. Biol.* **2,** 49–61.