STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **4**, September 2016, pp 233–242.
Published online in International Academic Press (www.IAPress.org)

A statistical model of macromolecules dynamics for Fluorescence Correlation Spectroscopy data analysis

Dmitri Koroliouk [1,*], Vladimir S.Koroliuk [2], Eleonora Nicolai [3], Paolo Bisegna [4], Lorenzo Stella [5], and Nicola Rosato [3]

[1]*Institute of Telecommunications and Global Information Space, National Academy of Sciences of Ukraine, Kiev;*
[2]*Institute of Mathematics, National Academy of Sciences of Ukraine, Kiev;*
[3]*Department of Experimental Medicine and Surgery, University of Rome Tor Vergata, 00133 Rome, Italy;*
[4]*Department of Civil Engineering, University of Rome Tor Vergata, 00133 Rome, Italy;*
[5]*Department of Science and Chemical Technology, University of Rome Tor Vergata.*

(Received: 24 March 2016; Accepted: 4 July 2016)

Abstract In this paper we propose a new mathematical model to describe the mechanisms of biological macromolecules interactions. Our model consists of a discrete stationary random sequence given by a solution of difference stochastic equation, characterized by a drift predictive component and by a diffusion term. The relative statistical estimations are very simple and effective, promising to be a good tool for mathematical description of collective biological reactions. This paper presents the mathematical model and its verification on simulated data set, obtained on the basis of the well-known Stokes-Einstein model. In particular we considered several mix of particles of different diffusion coefficient, respectively: $D1 = 10\mu m^2/sec$ and $D2 = 100\mu m^2/sec$. The parameters evaluated by this new mathematical model on simulated data, show good estimation accuracy, in comparison with the prior parameters used in the simulations. Furthermore, when analyzing the data for mix of particles with different diffusion coefficient, the proposed model parameters V (regression) and σ^2 (square variance of stochastic component) have a good discriminative ability for the molar fraction determination.

Keywords FCS, Brownian motion, Discrete Markov diffusion, Protein diffusion

AMS 2010 subject classifications 60J70, 62M05, 92C45

DOI: 10.19139/soic.v4i3.219

1. Introduction

Biological systems are intrinsically complex, due the huge number of interacting objects. In nanotechnology and nanobiotechnology applications, an important role is played by dynamics of macromolecules or interaction between them (proteins, metabolites, substrates,). In such wide view, mathematical models can be a useful tool for investigating a large number of questions in metabolism, biochemistry, genetics and gene-environment interactions, genotype-phenotype mapping and their use in medicine. Various complex systems like biochemical reactions are driven by binding / unbiding dynamics of the single species concentrations, described by mathematical models [18, 3, 7, 14].

Protein-ligand binding is a key biological process at molecular level. The identification and characterization of small-molecule binding sites on therapeutically relevant proteins could have tremendous implications for target

*Correspondence to: Dmitri Koroliouk (Email: dimitri.koroliouk@ukr.net). Department of Mathematical Modelling, Institute of Telecommunications and Global Information Space of the NAS of Ukraine, Kiev, Ukraine. Address: Ukraine, 03110, Kiev, Chokolovskii Bul'var, 13.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

evaluation and rational drug design [4]. Simulated statistical data based on simple chemical kinetic hypotheses can be used to validate the presence or absence of a binding in a set of molecules that makes up a complex system [11]. In some natural conditions and normalization, it has the asymptotical behavior as a discrete Markov diffusion [12]. In the simpler case of non-interacting species, the dynamic tracking of particles reduces to the study of mixtures of molecules with different speed of Brownian motion, due to their different hydrodynamic radii and to the corresponding different diffusion coefficients.

In the last two decades, the FCS technique has been developed, based on the laws of molecular diffusion, formulated from Browns observations of random particle motion and used to study the movements and the interactions of biomolecules at extremely dilute concentrations. FCS is a fluorescence based technique that can be used to study a variety of sample types. The analysis of FCS data examines the minute intensity fluctuations caused by spontaneous deviations from the mean at thermal equilibrium. These fluctuations can result from variations in local concentrations owing to molecular mobility or from characteristic intermolecular or intramolecular reactions of fluorescently labeled biomolecules present at low concentrations.

The analytical potential of FCS for the life sciences has been shown in numerous applications since the original work of Rigler and coworkers [1, 17] FCS was successfully used to study association and dissociation of nucleic acids [10] and proteins [15]. Moreover, these features allowed for real-time investigation of enzyme kinetics [9].

FCS is a method worth considering for a variety of biological and physicochemical questions. FCS has great advantages for quantitatively analyzing biomolecular mobilities and interactions in situ. Precise values of physical parameters, such as diffusion coefficients, are determined relatively easily and quickly. Essential information about processes governing molecular dynamics can, thus, be derived from the temporal pattern by which fluorescence fluctuations arise and decay. Although FCS curves are easy to record, interpreting them may not be straightforward. Many free parameters in the fitting procedure will return seemingly good fit results, even if the values for these parameters fail to reflect the sample properties or have no physical meaning at all. Any changes in molecular shape or size, on binding or cleavage, that affect the hydrodynamic radius of a particle, are also reflected in the diffusion coefficient and, thus, in the average diffusion time through the observation volume.

Interpretation of FCS data generally requires fitting the curves to a mathematical model. The type of applicable model and the parameters that can be extracted depends on the system under investigation. The diffusion coefficient D is derived from the characteristic decay time of the correlation function [19].

Usually, in order to estimate the diffusion coefficient and other parameters of interest, the autocorrelation function of the data is fitted typically using a nonlinear least squares algorithm. In this paper we introduce a new computational model and the corresponding statistical parameters characterizing the mobility of particles, in order to describe the dynamic of binding among different species. The model is based on difference stochastic equation with predictable linear component, and with an additional stochastic component, determined by a stationary, in the wide sense, random sequence. The statistical analysis of real measurements is particularly important when the observation of physical data (biological, chemical etc.) has often no justification in the form of simple physical theory.

In numerous complex systems like biological processes with equilibrium, the dynamics of concentration, or frequencies of some predefined attribute presence, can be described by the mathematical model of binary discrete Markov diffusions, based on statistical data of elementary hypotheses validation about the presence or absence of the predefined attribute in the set of elements that make up a complex system [11].

In our consideration, we proceed from the following principles of statistical analysis of physical data:

- all the elements that make up the system can gain or lose said attribute over time, that is the relative frequency of attribute, a discrete dynamic variable;
- the basic objects of our study are discrete Markov diffusions, characterized by relative frequencies of presence or absence of the attribute in a sample of fixed volume at each time instant;
- it is assumed that the average frequency of the attribute at a fixed time depends only on the average frequency of the attribute relative to the next previous time. Such relationship is called persistent regression and it is used as a fundamental condition for the subsequent analysis of the model.

The present work introduces a new computational method and the corresponding statistical estimators of parameters characterizing the dynamics of frequency (concentration). We proceed from the following

considerations: two different models, the physical (Stokes-Einstein) diffusion process and the mathematical one (stationary discrete Markov diffusion), describing the same physical process.

2. Mathematical model

2.1. Motivation of the model

The FCS method of macromolecules dynamic monitoring, registres the number of fluctuations of fluorescently labeled molecules in a confocal observation volume. The basic object of analysis is the fluctuation α_t of the fluorescence intensity, that is the difference of its current value at the instant $t \geq 0$ and its average value. The fluorescence intensity is proportional to the number of the labeled molecules observed at the instant t. A single molecule can freely diffuse in and out the observation volume or undergo chemical reactions. Thus resulting in a combination of motions of different species characterized by different drift and diffusion parameters. The mathematical model of stationary Gaussian diffusion, defined by a continuous process of Ornstein-Uhlenbeck type, is motivated by its important special case of motion of particles in a viscose liquid medium, known in physics as the Langevin equation. Consider the equation:

$$m\dot{v}(t) = -\beta v(t) + \mu_t \ , \ \ t \geq 0,$$

where m is the particle mass, $v(t)$ is the particle speed, β is the medium viscosity and μ_t is a noise component which will further be clarified.

Here, the component of the "force" $m\dot{v}(t)$ which slows down the particle, is proportional to the "speed" $v(t)$, and the presence of "noise" is associated with chaotic collisions of the particle with the molecules of the medium (due to thermal motion of the latter). In this formulation as the noise is used the random process $\mu_t = \dot{W}_t$, where $W = \{W_t \ , \ t0\}$ be a stochastic process of Brownian motion. According to Paley-Wiener-Zygmund Theorem, the random function W is nowhere differentiable almost surely.

In order to give meaning to the equation:

$$m\dot{v}(t) = -\beta v(t) + \dot{W}_t \ , \ \ t \geq 0, \tag{1}$$

known in physics as the Langevin equation, should be given a probabilistic meaning of the derivative process $\dot{W}_t$,$t \geq 0$. This is done by operating the concept of Ito stochastic integral.

2.2. Statistical model of stationary discrete Markov diffusion

In our case, we deal with discrete time, so using the structures like (1), we have to operate with finite (stochastic) differences. Therefore, we have no need to operate with stochastic integrals. In [11], a model of discrete Markov diffusions with persistent regression is considered as a stationary discrete Markov diffusion α_t, $t \in N_0 = \{0, 1, 2, ...\}$, given by a solution of the difference stochastic equation:

$$\Delta\alpha_t = -V\alpha_t + \sigma w_{t+1} \ , \ \ t \in \mathcal{N}_0 = 0, 1, 2, 3, ... \tag{2}$$

where

$$\Delta\alpha_t := \alpha_{t+1} - \alpha_t,$$

and

$$Ew_t = 0 \ , \ \ E(w_t)^2 = \sigma^2 \ , \ \ \forall t = 1, 2, ... \tag{3}$$

The stationary, in wide sense, random sequence α_t, $t = 0, 1, 2,$ is characterized by two numerical parameters:
V : regression parameter of the predictable component;
σ^2: variance of the stochastic component $\sigma w_(t + 1), t = 0, 1, 2,$ defined by a sequence of independent, identically distributed Gaussian random variables with parameters (3) or alternatively, by martingale-differences which satisfy the conditions of the Central Limit Theorem [8].

An important role is played by the condition of statistical stationarity for the discrete stationary discrete Markov diffusion [12]:

$$\sigma^2 = \sigma_0^2 V(2-V)\ ,\ \ 0 < V < 2, \tag{4}$$

where

$$\sigma_0^2 := cov(\alpha_t,\, \alpha_t) = E\alpha_t^2\ ,\ \ t \geq 0.$$

In this case, one can use several sufficient statistics which allow to estimate the numerical parameters of the model generated by the solutions of the difference stochastic equations (2) (3). In statistical analysis of stationary discrete Markov diffusion with persistent linear regression (2) (3), the main problem is to construct the statistical estimations of numerical parameters V (the regression) and σ^2 (the square variance of the stochastic component). The condition of statistical stationarity allows the use of the correlation analysis of the stationary discrete Markov diffusion. The covariances related to the stationary discrete Markov diffusion, defined by solution of equation (2) (3), in terms of the statistical stationarity (4), has the following form:

$$R := \sigma_0^2\ ,\ \ R^0 := E[\alpha_t \Delta\alpha_t]\ ,\ \ R^\Delta := E[\Delta\alpha_t \Delta\alpha_t]\ ,\ \ R^\Delta = R^\Delta(0) := 2V_0\sigma_0^2\ ,\ \ t \geq 0. \tag{5}$$

The following relations take place:

$$R^0 = -V\sigma_0^2\ ,\ \ R^\Delta = 2V\sigma_0^2 \tag{6}$$

and also

$$cov(\alpha_t,\, \alpha_{t+s}) = \sigma_0^2(1-V)^s\ ,\ \ s,t \geq 0\ ,\ \ 0 < V < 1. \tag{7}$$

The covariances (5) generate, as estimators, the following statistics:

$$R_T := \frac{1}{T}\sum_{t=0}^{T-1} \alpha_t^2\ ,\ \ R_T^0 = \frac{1}{T}\sum_{t=0}^{T-1} \alpha_t \Delta\alpha_t\ ,\ \ R_T^\Delta := \frac{1}{T}\sum_{t=0}^{T-1} (\Delta\alpha_t)^2 \tag{8}$$

The statistics (8) generate the following two estimators of the predictable component parameter V:

$$V \approx V_T^0 = -R_T^0/R_T\ ,\ \ V \approx V_T^\Delta = R_T^\Delta/2R_T, \tag{9}$$

which are strongly consistent and asymptotically unbiased.

Naturally one has the following obvious estimation

$$\sigma_0^2 \approx \sigma_{0T}^2 = R_T. \tag{10}$$

The stationarity condition (4) generates the stochastic component dispersion σ^2:

$$\sigma^2 \approx \sigma_T^2 = \mathcal{E}_T^0 R_T, \tag{11}$$

where

$$\mathcal{E}_T^0 := V_T^0(2 - V_T^0). \tag{12}$$

Remark 1

The properties of the statistical estimator V_T^0 are studied in [12]. The dispersion-type estimator V_T does not depend on its mean value.

Remark 2

The formula (7), under the condition $0 < V < 1$, can be also used as control of dynamics as follows:

$$\ln\, cov(\alpha_t,\, \alpha_{t+s}) = \ln\sigma_0^2 + s\ln(1-V)\ ,\ \ s,t \geq 0. \tag{13}$$

So in logarithmic scale, the covariance of the stationary discrete Markov diffusion is characterized by the linear function $a - sb$ where, by definition,

$$a := \ln\sigma_0^2\ ,\ \ b := -\ln(1-V). \tag{14}$$

The verification of stationary discrete Markov diffusion property can be done using the following estimations of the parameters a and b:

$$a \approx a_T = \ln R_T\ ,\ \ b \approx b_T = -\ln(1-V). \tag{15}$$

2.3. Kinetic Stokes-Einstein diffusion model

The discretized three-dimensional Brownian motions of n particles moving inside a cubic box of side L centered at the origin of a Cartesian frame have been obtained by numerical simulation (see also [3]). Each Cartesian component of the position of each particle has been modeled as an independent discretized scalar Brownian motion. Let W_t denote the value of such scalar Brownian motion at time t=$0, 1, 2$. The following time-marching scheme has been adopted:

$$W_t = W_{t-1} + dW_t \ , \quad t = 0, 1, 2, ... \tag{16}$$

where each dW_t is an independent random variable of the form $\sqrt{2D\delta}N(0,1)$. Here $N(0,1)$ denotes the normal random variable, δ is the sampling time interval, and D is the diffusion coefficient, given by

$$D = kT/6\pi r\eta \tag{17}$$

where kT is kinetic energy, r is radius of the particle, η is the viscosity of the medium. Periodic boundary conditions have been enforced at the box boundaries. The initial position of each particle was assumed to be a three-dimensional random variable, uniformly distributed inside the cubic box. A spherical region of interest (ROI) of diameter D_R centered at the origin has been considered. At each time instant t, the number α_t of particles inside the ROI has been recorded.

The algorithm has been implemented using massively parallel GPU computing and MATLAB.

Two simulation campaigns have been performed. First, all particles have been given the same diffusion coefficient D, and different values of D have been considered. Then, a mixture of two particle families with different diffusion coefficients D_1 and D_2 have been taken into account. The volume fraction of the slow-diffusing family was denoted by f.

The Parameter values adopted in the simulations are an ROI diameter of 1 μm, 10 particles expected in ROI giving a density of 19.099 particles/μm^3. Assuming a sampling frequency of 50 kHz (acquisition time: δ = 20 μs) with a simulation time of $10s$ (500 ksamples). The first run of simulation was performed setting the diffusion coefficient at different values, respectively: 10, 30, 50, 80, 100 $\mu m^2/s$. In the second run, two set of particles were considered to exist concurrently in the ROI, in detail: 10 and 100 $\mu m^2/s$ changing the ratio between the two species from 0 to 1, as schematized in the following table

Parameter	Description	Value/Range	Unit
D_R	ROI diameter	1	μm
L	side of cubic simulation box (periodic boundary conditions)	10	μm
N	total number of particles	19.099	$\sharp$
n_{ROI}	average number of expected particles in ROI	10	$\sharp$
P	average particle density	19.099	$\sharp/\mu m^3$
δ	sampling time interval	20	μs
f_S	sampling frequency	50	kHz
T	total simulation time	10	S
S	total number of samples	500,000	Samples
D	diffusion coefficient (first simulation campaign)	10, 30, 50, 80, 100	$\mu m^2/s$
$D_1,\ D_2$	diffusion coefficients (second simulation campaign)	10, 100	$\mu m^2/s$
F	volume fraction of the slow-diffusing family (second simulation campaign)	0, 0.1, 0.3, 0.5, 0.7, 0.9, 1	

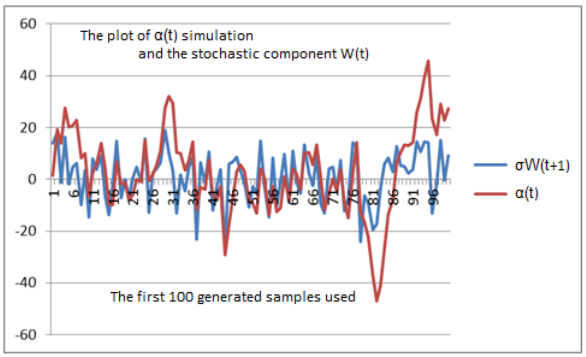
3. The stationary discrete Markov diffusion models verification by direct numerical simulation

The models verification by numerical simulation is performed by mean of numerical simulations of the trajectories of stationary discrete Markov diffusion (1) - (2) for specified parameters V and σ .

Next, according to statistical formulas (7) - (11), are calculated the values of estimates V_T^0, σ_{0T}^2 and σ_T^2 comparing with the original (theoretical) settings of parameters V, σ.

For different runs of the random numbers generator, we have the following result of the direct numerical simulation.

The following is the model numerical simulation and parameters estimation, case $V = 0,2$; $\sigma = 10$.

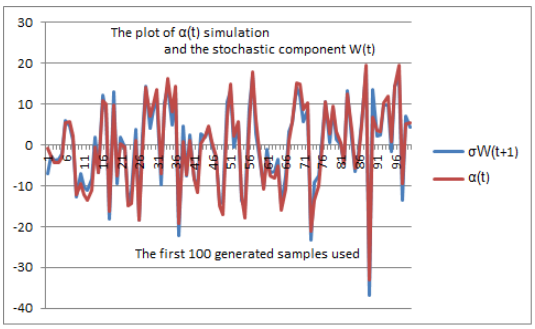


Δα(t+1)=-Vα(t)+σW(t+1)		V=0,2		σ=10	
PARAMETERS					
	Theor	Experimental evaluation			
		Run 1	Run2	Run 3	Run 4
σ	10	9,9878	9,978	10,057	10,0197
V	0,2	0,2066	0,201	0,1997	0,1989

Figure 1. The case (V=0,2; σ=10). Left: plot of a numerical simulatied discrete Markov diffusion $\alpha(t)$ and its stochastic component $\sigma W(t+1)$. Right: simulated discrete Markov diffusion parameters evaluation for 4 different runs of the random numbers generator.

In our calculations, we use 30000 samples of the data length, which give good convergence.

The model verification is done for a wide range of V ($0 < V < 1$) and σ ($\sigma > 0$), so we give a calculation for another original setting of parameters V, σ.



Δα(t+1)=-Vα(t)+σW(t+1)		V=0,8		σ=10	
PARAMETERS					
	Theor	Experimental evaluation			
		Run 1	Run2	Run 3	Run 4
σ	10	10,0237	9,951	10,012	10,0248
V	0,8	0,80173	0,792	0,8103	0,7977

Figure 2. The case V=0,8; σ=10. Left: plot of a numerical simulatied discrete Markov diffusion $\alpha(t)$ and its stochastic component $\sigma W(t+1)$. Right: simulated discrete Markov diffusion parameters evaluation for 4 different runs of the random numbers generator.

The discrete Markov diffusion numerical simulation gives a convinced statistical estimation of the parameters V, σ.

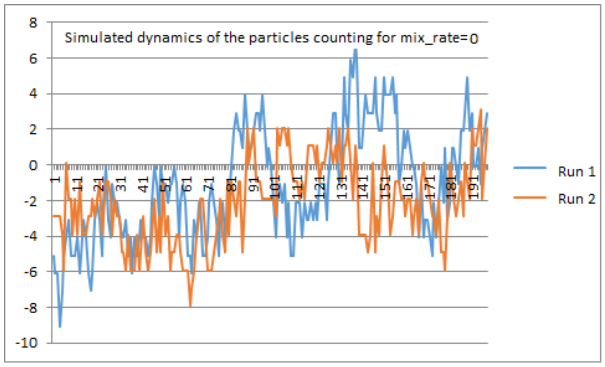
4. The stationary discrete Markov diffusion models verification using Kinetic Stokes-Einstein diffusion model simulated data

The data, based on Stokes-Einstein diffusion model, present the dynamics of the number of numerically simulated particles, counted in a limited volume in pre-set time slots (20μs), for the total observation time interval about 0,328 sec. (16385 samples). The simulation generates the mix of two Brownian motions: a slow one (D=10) and a fast one (D=100). The samples are obtained for different mix rate (proportions) of the fast and the slow particles with mix rate values: (0|1); (0,1|0,9); (0,3|0,7); (0,5|0,5); (0,7|0,3); (0,9|0,1) and (1|0). For every mix rate value, several samples are acquired by mean of different runs of the random numbers generator.

Now the 7 groups of the samples are considered as trajectories of stationary discrete Markov diffusion, defined by (2) - (3), with unknown parameters V, σ estimated by the statistics (8) - (12).

The use of stationary discrete Markov diffusion model in the analysis of statistical data, based on the Stokes-Einstein diffusion model, gives effective mechanisms of interactions analysis of slow and fast diffusions (see Fig. 3, 4). Let us first consider the two extreme cases mix-rate =(0|1) and (1|0).

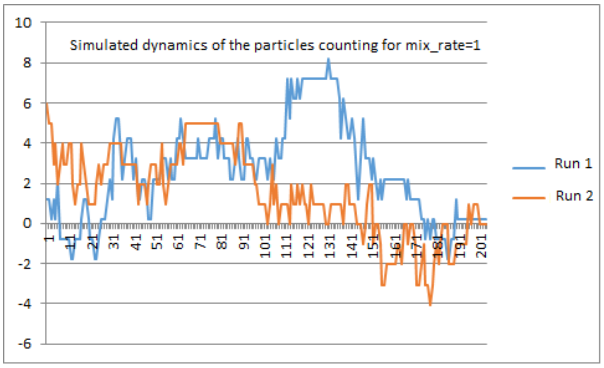
For the mix-rate =(0|1) one can obtain, using the estimators (8) - (12), the following parameters evaluations:



Parameter	Run 1	Run 2
$\sigma_0^2 \approx \Sigma\alpha^2(k)/T$	9,957267597	9,838731687
$V \approx \Sigma\Delta\alpha^2/2\Sigma\alpha^2$	0,151887047	0,151482925
$V \approx -\Sigma\alpha\Delta\alpha/\Sigma\alpha^2$	0,151887416	0,15148301
W=1–V/2	0,924056476	0,924258538
C=2V·W	0,280704419	0,280018773
$\sigma^2=C\cdot\sigma_0^2$	2,795049017	2,755029577
$\sigma=\sqrt{\sigma^2}$	1,671840009	1,659828177

Figure 3. The mix-rate =(0|1). Left: Stokes-Einstein kinetic model trajectory. Right: the stochastic discrete Markov diffusion model parameters evaluation.

The same calculations for the mix-rate = (1|0) give the following:



Parameter	Run 1	Run 2
$\sigma_0^2 \approx \Sigma\alpha^2(k)/T$	9,5117969	9,681002702
$V \approx \Sigma\Delta\alpha^2/2\Sigma\alpha^2$	0,049345364	0,049946949
$V \approx -\Sigma\alpha\Delta\alpha/\Sigma\alpha^2$	0,049345429	0,049947382
W=1–V/2	0,975327318	0,975026526
C=2V·W	0,096255763	0,0973992
$\sigma^2=C\cdot\sigma_0^2$	0,91556527	0,942921915
$\sigma=\sqrt{\sigma^2}$	0,956851749	0,971041665

Figure 4. The mix-rate =(1|0). Left: Stokes-Einstein kinetic model trajectory. Right: the stochastic discrete Markov diffusion model parameters evaluation.

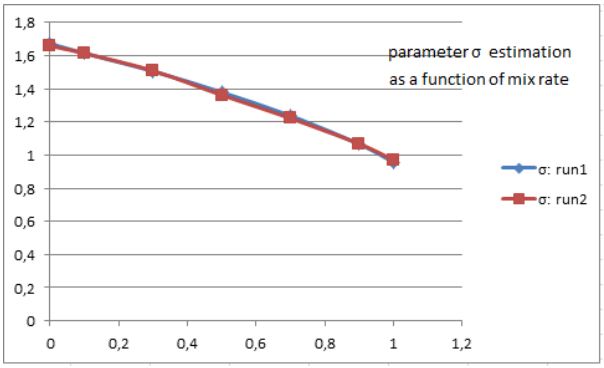
One can see that the values of the main parameters V, σ in both the cases are very different. This is an indication that the main parameters V, σ can also be discriminant for any value of the mix-rate. So now we should to repeat

the same calculus of the model parameters for the mix-rate values (0,1|0,9); (0,3|0,7); (0,5|0,5); (0,7|0,3), (0,9|0,1) and then to compare the dynamics of parameters V, σ and σ_0^2 by variating the mix-rate.

5. The mix rate discrimination capability of the fundamental parameters of the model using Kinetic Stokes-Einstein diffusion model simulated data

The simulated data, based on Stokes-Einstein diffusion model, has been produced for seven mix-rate values (0,1|0,9); (0,3|0,7); (0,5|0,5); (0,7|0,3) and (0,9|0,1). Every simulation run contains 500000 samples, used for evaluation of parameters V, σ and σ_0^2 by direct application of the estimators (8) - (12).

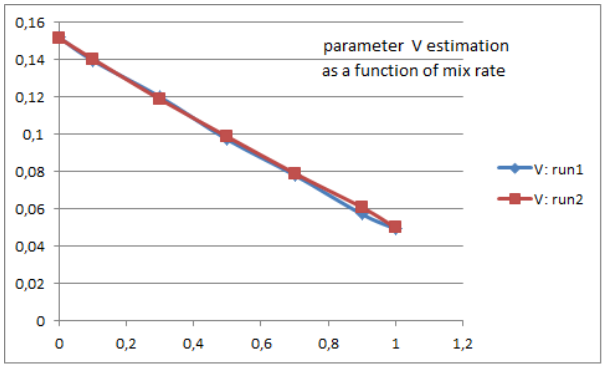
The following results has been obtained for the parameter σ :



mix_rate	σ: run1	σ: run2
0	1,671841171	1,65982983
0,1	1,614770647	1,61374915
0,3	1,505486209	1,50890114
0,5	1,379160156	1,36214922
0,7	1,237259001	1,2220825
0,9	1,066198549	1,06760628
1	0,956851749	0,97104166

Figure 5. Left: the plot of σ estimated values, as a function of mix-rate. Right: the parameter σ estimation for the 7 groups of the samples acquired by 2 different runs of the random numbers generator.

The parameter V evaluation gives us the following:



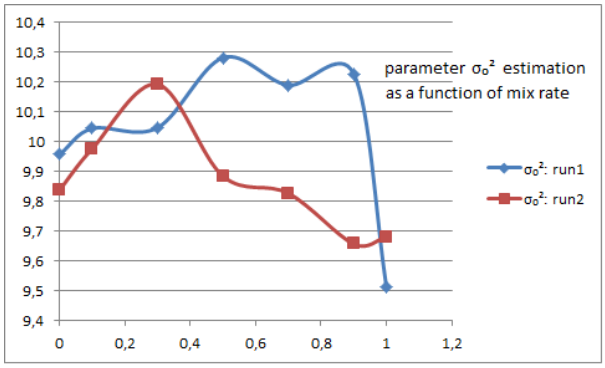
mix_rate	V: run1	V: run2
0	0,151888176	0,15148295
0,1	0,139527392	0,14036579
0,3	0,11998634	0,11873494
0,5	0,097219474	0,09873162
0,7	0,078169653	0,07911793
0,9	0,057210931	0,06086607
1	0,049345364	0,04994695

Figure 6. Left: the plot of V estimated values, as a function of mix-rate. Right: the parameter V estimation for the 7 groups of the samples acquired by 2 different runs of the random numbers generator.

And for the parameter σ_0^2 we have:

One can see from the tables and relative plots that, the parameters σ and V have good ability to distinguish the mix rate of two ensembles of the particles with different diffusion coefficient, according to the Stokes-Einstein diffusion model.

However, the parameter σ_0^2 does not have such property.



mix_rate	σ_0^2: run1	σ_0^2: run2
0	9,957213516	9,8387497
0,1	10,04474549	9,97661571
0,3	10,04756329	10,1927893
0,5	10,28223292	9,88438591
0,7	10,18985516	9,82710111
0,9	10,22754759	9,65693251
1	9,5117969	9,6810027

Figure 7. Left: the plot of σ_0^2 estimated values, as a function of mix-rate. Right: the parameter σ_0^2 estimation for the 7 groups of the samples acquired by 2 different runs of the random numbers generator.

6. Conclusions

The following results has been obtained:

1. A new statistical model, presented by the solution of equation (2) - (3) is proposed for the mathematical description of macromolecules interaction, which is characterized by numerical parameters V (regression) and σ^2 (the square variance of the stochastic component).
2. There were obtained statistical estimators (8), (11) and (12) of the interaction parameters (V, σ^2) promised to be far superior in accuracy and efficiency than the traditionally used method of autocorrelation curve fitting, based on kinetic Stokes-Einstein model (16) - (17).
3. The statistical estimators (8) - (12), applied to a numerically simulated process defined by the statistical model (2) - (3), give excellent coincidence of the parameters evaluation with their real value, which proves the high accuracy of estimators (8) - (12).
4. The proposed model has been tested and verified on a set of simulated data derived from a discrete kinetic Stokes-Einstein Brownian process of two biological molecules with different diffusion coefficients and different molar fractions. The statistics (8) - (12) has quasi-linear discriminant capacity for the proportion of different molar fractions containment.

Generally the biochemical systems are quite complex, involving many species and many reactions, thus it is difficult to study and analyze them. The strategy provided utilizes an experimental technique combined with stochastic difference equations to extrapolate parameters directly related to biochemical system dynamics. The proposed model may be used for a large class of biochemical reaction models. A possible application in biophysics is the study of biological macromolecules diffusion in solution (eg.: enzymes and substrates), combining our model with dataset acquired by a fluorescence correlation spectroscopy system. In this case, in fact, the fluorescent fluctuations signal of marked molecules has been observed in a reduced number of molecules in a small measuring volume (about 1 cubic micron) defined by the focusing of excitation laser. The speed of these fluctuations depends, in turn, from the diffusion coefficient of molecules under study, hence by its hydrodynamic radius. Applying in next works the proposed model to experimental data of fluorescence correlation spectroscopy, it will be possible to directly determine the diffusion coefficients and molar fractions of molecules under study. In particular, it will be possible to follow the dynamic of binding/unbinding of enzymatic macromolecules and its substrate, accordingly to the mass variation of the observables. These processes could be of fundamental importance for the development of new drugs as inhibitors of artificial enzymes involved in pathologies. To understand complex biological systems, it is necessary to obtain a thorough understanding of the interaction between molecules and pathways. Such fact is even truer for understanding complex diseases such as cancer, Alzheimer and others as well. For these reasons, we believe the method here presented could provide a new important tool to improve biomedical and pharmaceutical research.

Acknowledgement

One of the authors, D. Koroliouk, was supported by the Tor Vergata University Visiting Professorship Program in 2014 which provided good opportunities for interdisciplinary collaboration in a broad international team.

REFERENCES

1. M. Eigen, and R. Rigler, *Sorting single molecules: application to diagnostics and evolutionary biotechnology*, Proceedings of the National Academy of Sciences, vol. 91, pp. 5740–5747, 1994.
2. L. Fay, and A. Balogh, *Determination of reaction order and rate constants on the basis of the parameter estimation of differential equations*, Acta Chimica Academiae Scientarium Hungarica, Budapest, vol. 57, 1968.
3. H. P. Fischer, *Mathematical Modeling of Complex Biological Systems*, Alcohol Research & Health, vol. 31, pp. 49–59, 2008.
4. Z. Guo, B. Li, L. T. Cheng, S. Zhou, J. A. McCammon, and J. Che, *Identification of Protein-Ligand Binding Sites by the Level-Set Variational Implicit-Solvent Approach*, Journal of Chemical Theory and Computation, vol. 11, pp. 753–765, 2015.
5. D. J. Higham, *An algorithmic introduction to numerical simulation of stochastic differential equations*, Society for Industrial and Applied Mathematics Review, vol. 43, pp. 525–546, 2001.
6. L. H. Hosten, *A comparative study of short cut procedures for parameter estimation in differential equations*, Computers and Chemical Engineering, vol. 3, 1979.
7. S. Ilie, S. Gholami *Simplifying Stochastic Mathematical Models of Biochemical Systems*, Applied mathematics, vol. 4, pp. 248–256, 2013.
8. J. Jacod, A. N. Shiryaev *Limit Theorems for Stochastic Processes*, Springer Berlin - Heidelberg - New York, 2003.
9. U. Kettling, A. Koltermann, P. Schwille, R. Eigen *Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy*, Proceedings of the National Academy of Sciences vol. 95, pp. 1416–1420, 1998.
10. M. Kinjo, R. Rigler *Ultrasensitive hybridization analysis using fluorescence correlation spectroscopy*, Nucleic Acids Research vol. 23, pp. 1795–1799, 1995.
11. D. Koroliouk, V. S. Koroliuk, N. Rosato *Equilibrium Process in Biomedical Data Analysis: the Wright-Fisher Model*, Cybernetics and System Analysis, Springer NY, vol. 50, no. 6, pp. 890–897, 2014. DOI: 10.1007/s10559-014-9680-y.
12. D. Koroliouk *Stationary statistical experiments and the optimal estimator for a predictable component. Journal of Mathematical Sciences*, Springer NY, vol. 214, no. 2, , pp. 220–228, 2016. DOI: 10.1007/s10958-016-2770-9.
13. A. V. Karnaukhov, E. V. Karnaukhova, J. R. Williamson *Numerical Matrices Method for Nonlinear System Identification and Description of Dynamics of Biochemical Reaction Networks*, Biophysical Journal, vol. 92, pp. 3459–3473, 2007.
14. S. Rao, A. Van Der Schaft, K. Van Eunen, B. M. Bakker, B. Jayawardhana *A model reduction method for biochemical reaction networks*, BMC Systems Biology, vol. 8, no. 52, 2014.
15. B. Rauer, E. Neumann, J. Widengren, R. Rogler *Fluorescence correlation spectrometry of the interaction kinetics of tetramethylrhodamin alpha-bungarotoxin with Torpedo californica acetylcholine receptor*, Biophysical Chemistry, vol. 58, pp. 3–12, 1996.
16. S. Schuster, C. Hilgetag, J. H. Woods, D. A. Fell *Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism*, Journal of Mathematical Biology, vol. 45, pp. 153–181, 2002.
17. P. Schwille, J. Bieschke, F. Oehlenschlager *Kinetic investigations by fuorescence correlation spectroscopy: the analytical and diagnostic potential of diffusion studies*, Biophysical Chemistry, vol. 66, pp. 211–228, 1997.
18. A. V. Skorokhod, F. C. Hoppensteadt, H. Salehi *Random Perturbation Methods with Applications in Science and Engineering*, Springer AMS, New York, vol. 150, 2002.
19. P. Schwille, F. J. Meyer-Almes, R. Rigler *Dual-color fluorescence cross-correlation spectroscopy for multicomponent diffusional analysis in solution*, Biophysical Journal, vol. 72, no. 4, pp. 1878–1886, 1997.
20. S. Vajda, P. Valko, A. Yermakova *A direct-indirect procedure for estimating kinetic parameters*, Computers and Chemical Engineering, vol. 10, pp. 49–58, 1986.
21. W. Vance, A. Arkin, J. Ross *Determination of causal connectivities of species in reaction networks*, Proceedings of the National Academy of Sciences of the United States of America, vol. 99, pp. 5816-5821, 2001.