Stochastic Arithmetic as a Tool to Study the Stability of Biological Models

Rene Alt

Laboratoire d'Informatique de Paris 6
University Pierre et Marie Curie, 4 place Jussieu,
75252 Paris cedex 05, France
e-mail: Rene.Alt@lip6.fr

Jean-Luc Lamotte

Laboratoire d'Informatique de Paris 6
University Pierre et Marie Curie, 4 place Jussieu,
75252 Paris cedex 05, France
e-mail: Jean-Luc.Lamotte@lip6.fr

Abstract—**The theoretical study of the stability of the numerical solution of a differential system may be complicated or even not feasible when the system is large and nonlinear. Here it is shown that such a study can be experimentally done by using stochastic arithmetic and its discrete approach known as the CESTAC method.**

The CESTAC method has been first proposed since more than forty years by M. La Porte and J. Vignes as an experimental statistical method to estimate the accuracy on the result of numerical program [10], [14]. Later an abstract formalization of the theory called Stochastic Arithmetic has been developed and many of its algebraic properties have been studied [2], [4], [5]. Here a brief presentation of stochastic arithmetic, of its main properties and of the different software existing for its implementation are given. Then it is demonstrated that the use of stochastic arithmetic in the solver of a differential system can easily reveal whether the computed solution is stable or not. Moreover the stability can be studied with respect to the coefficients of the system or with respect to the initial conditions. At the end it is also pointed out that the same method can be used to detect instabilities due to the used solver. Some examples taken from the biological literature are given [1], [6], [7].

Keywords-**Stochastic arithmetic, CESTAC method, stability of differential biological models.**

I. INTRODUCTION

The detection of instabilities in the numerical solution of differential systems is generally not an easy thing to do. Actually instabilities have two main sources. The theoretical differential system can be stiff or inherently unstable and the numerical method used to solve it can also be unstable. A common example of this last situation is the numerical solution of a stiff system using an explicit method and a too large step.

The classical approach to know whether a differential system is stable or not is the computation of its Jacobian and of its eigenvalues, see for example [12]. This is generally not so easy and requires most of the time manual calculation or the use of a computer algebra system. In the same manner the use of an explicit method to solve a differential system is rather simple but may lead to numerical instabilities if the step happen to be too large even if the method has an automatic calculation of the step. On the contrary an implicit method may not have this step restriction but requires at each step the solution of a system of equations which is non linear when the differential system is non linear. And this is the case for most biological models. Here it is shown that a simple method called the CESTAC method can be easily used to investigate the sensitivity of the computed solution to the coefficients and initial conditions of a differential system and to detect possible instabilities. Various numerical examples coming from the modelisation of bacterial growth are presented.

The structure of this paper is as follows. First stochastic arithmetic and the CESTAC method are shortly recalled. Then a software called CADNA which imple-

ments the CESTAC method is presented and the use of this software to investigate the stability of the solution of a differential system is detailed. Its efficiency is illustrated on several biological models.

II. Stochastic arithmetic and the CESTAC method

A. Stochastic Arithmetic

Stochastic arithmetic is a model for exact computation on imprecise data. It can be summed-up as follows.

Let us assume that an imprecise data can be represented as a Gaussian distribution with known mean value m and known standard deviation σ, $\sigma \geq 0$. In the following such an imprecise data is called a stochastic number. Thus the set of stochastic numbers denoted $\mathbb{S}$ is the set of Gaussian random variables.

One of the main properties of a Gaussian distribution and hence of a stochastic number is:

For $X = (m, \sigma)$ the confidence interval of m with a probability $\mathcal{P} = (1 - \eta)$ is defined as $[m - \lambda_\eta \sigma, m + \lambda_\eta \sigma]$ where λ_η is such that

$$\mathcal{P}\left(X \in [m - \lambda_\eta \sigma, m + \lambda_\eta \sigma]\right) = 1 - \eta, \qquad (1)$$

It is well known that for $\eta = 0.05$, that is $\mathcal{P} = 1 - \eta = 0.95$, we have $\lambda_\eta = 1.96$. Consequently the number of significant decimal digits on m is the integer part of:

$$C_{\eta,X} = log_{10}\left(\frac{|m|}{\lambda_\eta \sigma}\right), \qquad (2)$$

provided that $|m|/(\lambda_\eta \sigma) \geq 10$, otherwise we assume $C_{\eta,X} = 0$.

The ratio $\lambda_\eta \sigma / |m|$ is called the relative error of the stochastic number X. This characteristic is analogous to the relative error of an approximate number.

The arithmetic operations on stochastic numbers are defined as the operations on independent Gaussian distributions. They are denoted $_s+$, $_s-$, $_s*$, $_s/$ in [8] and [9] but here the simpler notations $+, -, \times, /$ are preferred. They are:

$$X_1 + X_2 \overset{def}{=} \left(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$
$$X_1 - X_2 \overset{def}{=} \left(m_1 - m_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$$
$$X_1 \times X_2 \overset{def}{=} \left(m_1 m_2, \sqrt{m_2^2 \sigma_1^2 + m_1^2 \sigma_2^2 + \sigma_1^2 \sigma_2^2}\right)$$
$$X_1 / X_2 \overset{def}{=} \left(m_1/m_2, \sqrt{\left(\frac{\sigma_1}{m_2}\right)^2 + \left(\frac{m_1 \sigma_2}{m_2^2}\right)^2 + \left(\frac{\sigma_1 \sigma_2}{m_2^2}\right)^2}\right)$$
$$\qquad (3)$$

The first three formulae including stochastic multiplication are exact. The formula for the division is only exact to the first order terms in σ/m and must be considered as an approximation. Actually it is well known that the distribution of the ratio of two Gaussian variables with expected value 0 and variance 1 is not Gaussian but follows a Cauchy law which has no mean value and no standard deviation but is symmetric and has a mode and a median. More details on stochastic arithmetic can be found in [2], [4], [5], [11], [13].

B. The CESTAC method

When one wants to develop a software to estimate the accuracy of a numerically computed result a first possibility is to use formulae (3) instead of standard floating point operations. This can be easily done as many programming languages such as C++ or Fortran 90 allow the overloading (re-definition) of the floating point operations.

Another approach used in the CESTAC method, see [3], [14], [9] is to discretize the theoretical Gaussian distributions with Gaussian random samples and to use their empirical mean values and standard deviations instead of the theoretical ones. This is done in the CESTAC method in the case of rounding errors coming from the floating point operators.

The idea of the CESTAC method is that each result of a floating point operator (assignment, arithmetic operator) which is not an exact floating point value, is always bounded by two floating point values R^- and R^+ obtained by rounding up or down the exact result, each of them being representative of the exact result. The random rounding mode consists, at the level of each floating point operation or assignment, in choosing as a result, randomly with an equal probability, either R^- or R^+. Thus when a code is performed N times in a synchronous parallel way with the use of this random rounding mode, N samples $R_k, k = 1..., N$ of each computed results are obtained, and then from these samples, the accuracy of the mean value $\overline{R}$ of these samples, considered as the computed result, may be estimated.

Hence a probabilistic model of the round-off error on a computed result obtained with the random rounding mode has been developed, see [9]. In this model it is shown that under two simple hypotheses which generally hold in real life problems, each sample obtained by the CESTAC method may be modelled by a random variable

Z defined by:

$$Z \simeq r + \sum_{i=1}^{nb} u_i(d) 2^{-p} z_i, \qquad z_i \in\,]-1, +1[\qquad (4)$$

where r is the unknown exact result, p is the length of the mantissa in the floating point representation of numbers, $u_i(d)$ are constants, nb is the number of arithmetic operations and z_i are independent centred and equi-distributed variables.

As a consequence: $E(Z) \simeq r$, the distribution of Z is quasi-Gaussian and the estimation of the accuracy on $\overline{R}$ can be done with the Student test which provides a confidence interval for $\overline{R}$.

Hence the number of decimal significant digits $C_{\overline{R}}$ of $\overline{R}$ can be estimated by:

$$C_{\overline{R}} = log_{10}\left(\frac{\sqrt{N}\,|\overline{R}|}{\tau_\eta \sigma}\right) \qquad (5)$$

where

$$\overline{R} = \frac{1}{N}\sum_{k=1}^{N} R_k, \qquad \sigma^2 = \frac{1}{N-1}\,\left(R_k - \overline{R}\right)^2$$

τ_η is the value of the Student distribution for $N-1$ degrees of freedom and a probability level $1-\eta$.

From a theoretical point of view the vector of the N empirical values representing a floating point result used in the CESTAC method is called a discrete stochastic number. In the same manner the CESTAC method is said to use a discrete stochastic arithmetic (DSA).

It must be noted that the CESTAC method differs from a simple Monte Carlo method where a Gaussian noise would be added to the data and the program would be run several times. In contrast in the CESTAC method a Gaussian noise is actually added to the data but also after each arithmetic operation N results are computed being rounded randomly up or down. Moreover the runs are done in such a way that at each test the same branching is performed in all runs. Thus after each arithmetic operation or each test the theoretically computed stochastic number appears as a vector of N empirical values really representing the same theoretical value. Hence, the number of decimal significant digits of any intermediate result can be computed using formula (5) in the same manner as the one(s) of the final result(s). This would not be the case in a simple Monte Carlo method.

C. The Cadna software

The Cadna software implements the so-called discrete stochastic arithmetic, which is nothing else than the CESTAC method to which have been added comparison operators, the notion of non-significant result and some more complementary features. It can be freely down-loaded from [8]. Two versions exist, one in Fortran 90 and one in C++. They have been developed as libraries to be added to an already existing code.

In this software, new types for single precision and double precision stochastic numbers have been defined and all arithmetic operators and tests have been over-loaded so that computing with stochastic numbers is as easy as computing with real numbers. Thus, any Fortran 90 or C++ code working with real numbers can be almost instantly converted in a code working with stochastic numbers, i.e. numbers with their errors. It has been theoretically and experimentally shown that formula (5) is correct to one digit with $N = 3$, consequently in the Cadna software all stochastic numbers are represented as three samples with Gaussian distribution.

Another feature of the Cadna software is that stochastic numbers represent imprecise numbers where the error is due not only to rounding in floating point arithmetic but also to the data which may also be imprecise. Thus in this software it is possible to introduce errors in the data so that an imprecise data is represented by a stochastic number with a known mean value and a known standard deviation. This possibility is used in the experimental investigation of the stability of the numerical solution of differential systems especially those coming from the modelisation of biological reactions.

III. APPLICATION TO SOME BIOLOGICAL MODELS

Many biological models are represented as nonlinear differential systems. Studying their stability may be difficult, see for example [6]. But the use of the Cadna software leads to an immediate answer to the question: "Is the computed solution stable around this special value of this particular parameter?"

Of course this is not the answer to the more general question "What is the domain of stability of the system?" but as is shown below, it can easily help to analyse the sensitivity of the solution to some parameter or to initial conditions.

As illustrations of the efficiency of the CESTAC method to experimentally investigate the stability of computed solutions, several biological model solutions have been computed using a fourth order Runge-Kutta method together with the Cadna software.

Errors have been introduced in the initial conditions and in some parameters to see the effect of imprecise data on the computed solution. The results are reported in the corresponding figures. In these figures the solutions computed with discrete stochastic numbers are represented as three curves corresponding to the three samples as explained above.

A. First model

The model is taken from [7] for cyclodextrin-glucano-transferase production by immobilised cells of Bacillus circulans ATCC21783. The differential system is:

$$\begin{array}{ll}
\frac{dx}{dt} & = (\mu(s) - \gamma)x \\
\frac{ds}{dt} & = -\mu(s)\,\frac{x}{y_{x/s}} \\
\frac{dp}{dt} & = (\alpha\mu(s) + \beta)x
\end{array} \qquad (6)$$

with initial conditions s_0, x_0, p_0. In this system x is the biomass concentration, s is the substrate concentration and p is the product concentration. $\mu(s)$ is the Andrews function: $\mu(s) = \mu_{max}\,\frac{s}{K_s + s + K_i s^2}$.

The experiments have been done with the coefficients: $\alpha = 1.11, \beta = 0.07, \gamma = 0.06, \mu_{max} = 2, y_{x/s} = 26.3,\ K_s = 0.8,\ K_i = 0.12,\ s_0 = 2,\ x_0 = 0.2,\ p_0 = 0$

This system has been successively solved with some tolerances introduced on the initial condition for the substrate ($s_0 = 2 \pm 0.2$), then for the biomass concentration ($x_0 = 0.2 \pm 0.1$), then on the coefficient μ_{max} ($\mu_{max} = 2 \pm 0.3$) and on the coefficient $y_{x/s}$ ($y_{x/s} = 26 \pm 1$). At last the system has been solved using a too large step ($h = 0.9$) so that the numerical integrating method is unstable.

The results are reported in figures 1 to 5.

A simple observation of these figures shows that errors on the initial conditions or on the coefficients cause two kinds of modifications on the solution. One kind is an increase or decrease of the limit of the product as in figures (1) and (4), the other kind is a modification of the delay after which the biomass and product begin a fast increasing as in figures (2) and (3). Another conclusion is that the system is much more sensitive to the initial substrate concentration than to the initial biomass concentration.

B. Second model

The model is taken from Alt and Markov [1] for E. Coli + Glucose.

In this model the microbial population is subdivided into two subgroups:

i) micro-organisms in lag and stationary phase are classified into one subclass with biomass denoted x. It is assumed that micro-organisms in that class experience unfavourable growth conditions and are not able to immediately produce anything;

ii) active (viable) micro-organisms in log phase, denoted y, possessing a complete set of active enzymes.

Bacteria in dying state are modelled by decay terms and need not be assigned to a special subgroup.

The system of equations is here:

$$\begin{array}{ll}
ds/dt & = -k_1 xs - \beta ys, \\
dx/dt & = -k_1 xs + k_2 y - k_d x^2, \qquad (7)\\
dy/dt & = k_1 xs - k_2 y + \beta ys,
\end{array}$$

with the initial conditions

$$s(0) = s_0,\ x(0) = x_0,\ y(0) = y_0.$$

The terms participating in this system have the following meaning:

$k_1 xs$ models the consumption of s by bacteria x and the transition of (fasting) bacteria x into (viable, active) bacteria y; βys models the consume of s by bacteria y and the increase of bacteria biomass y due to nutrition and reproduction; $k_2 y$ models the random transitions of bacteria from class y into class x; $k_d x^2$ models competition and decay of (starving) bacteria x.

This system has been solved with the following coefficients and initial conditions:

$k_1 = 0.23,\ k_2 = 0.85,\ k_d = 0.3,\ \beta = 1.0,\ s_0 = 2.0,\ x_0 = 0.25,\ y_0 = 0$.

To study the stability of the solution with respect to the initial conditions, some relative errors have been successively introduced into them. They were: 5% on s_0 and 25% on x_0. The results are given in figure (6) and (7). These two figures clearly show that the solution is much more sensitive to an error on the initial amount of substrate than to the initial amount of biomass. In fact a 25% relative error on the biomass perturbs the solution less that a 5% error on the initial substrate. This remark goes in the same direction as the corresponding one for the first model.

C. Third model

The model is taken from [6] for 1,2-dichloroethane (DCA) biodegradation by Klebsiella oxytoca va 8391 immobilized on granulated carbon. The differential system

is:

$$dx_1/dt = (\mu_1(s) - D)x_1 + k_{im}x_{im}$$
$$dx_{im}/dt = (\mu_{im}(s) - k_{im})x_{im}$$
$$ds/dt = -(\tfrac{1}{\gamma}\mu_1(s) + \beta_1)x_1 - (\tfrac{1}{\gamma}\mu_{im}(s) + \beta_{im})x_{im}$$
$$+D(s^{in} - s) - k_{La}(1 - \mu_2(s))s$$
$$dp/dt = (\tfrac{1}{\gamma}\mu_1(s) + \beta_1)x_1$$
$$+(\tfrac{1}{\gamma}\mu_{im}(s) + \beta_{im})x_{im} - Dp$$

$$(8)$$

with:

$$
\begin{aligned}
\mu_1(s) &= \frac{m_1 s}{k_s + s + s^2/k_i} \\
\mu_{im}(s) &= \frac{m_{im} s}{k_s + s + s^2/k_i} \\
\mu_2(s) &= \frac{m_2 s}{k + s}
\end{aligned}
\qquad (9)
$$

with initial conditions $x_1(0)$, $x_{im}(0)$, $s(0)$, $p(0)$.

The signification and values of the coefficients of this system can be found in [6]. This model differs from the classical bioreactor models as a phase with immobilised cells has been added to the normal free cells. These cells are attached to carrier particles and can grow and detach from the solid surface to leak into the liquid. In this model x_1, x_{im}, s, p respectively represent the free cells, the immobilised cells, the substrate and the product. A detailed theoretical study of the stability of this model has been developed in the cited paper.

Two experiments are reported here. The solutions have been computed with the coefficients proposed by the authors and initial conditions: $x_1(0) = 0.02$, $x_{im}(0) = 9$, $s(0) = 0.25$, $p(0) = 0$, which satisfy the theoretical stability conditions.

In the first experiment relative errors have been simultaneously introduced in the initial conditions: 10% on $x_1(0)$, 50% on $x_{im}(0)$ and 40% on $s(0)$, i.e. the initial values have been N times ($N = 3$) randomly generated in the intervals $x_1(0) \pm 10\%$, $x_{im}(0) \pm 50\%$, $s(0) \pm 40\%$.

The second experiment has been done with the same coefficients but with 100% simultaneous relative errors in the initial conditions $s(0)$, $x_1(0)$ and $x_{im}(0)$. The results are shown in figures (8) and (9).

It must be noted that the concentrations of free cells and of immobilized cells are of different order, close to $0.3\ kg/m^3$ for the free cells and close to $200\ kg/m^3$ for the immobilised cells. This is why in the figures the concentrations of immobilised cells have been scaled by a factor 0.001.

In the second experiment the Cadna software detects several instabilities as some multiplications and divisions have non-significant results. Anyhow the corresponding figure (9) shows that the asymptotes of all components do not depend on the initial conditions. Indeed this fact

which can probably be proved has been numerically experimented by running the program with many different initial conditions and the asymptotes are always identical provided that $x_{im}(0) \neq 0$ and that $s(0)$ satisfies the stability condition:

$$s_0 \leq Ds^{in} - \beta_1 x_1(0) - \beta_{im} x_{im}(0)$$

as explained in [6].

IV. CONCLUSION

Stochastic arithmetic has been proved to be an interesting method for the estimation of the error on a computed result when the data are inaccurate and the arithmetic operators introduce round-off errors. In this paper it has been experimentally shown on several bioreactor models that the CESTAC method and the corresponding Cadna software which are based on stochastic arithmetic can provide an easy alternative to theoretical studies when one wants to know whether the computed solution of a differential system is correct and stable or not. Of course the Cadna software does not lead to the domain of stability of the system but only to the knowledge of the stability of a particular solution computed with a particular numerical method. But in many cases the interest is not in the whole domain of stability but only in the solution of a particular model of a real experimental bioreactor. As presented on the models taken from the classical literature, the Cadna software is particularly efficient in the detection of the coefficients or initial values, a small variation of which introduces a large variation in the solution. In the same idea, the coefficients which have a very little influence on the solution can be detected as well. This gives a precious information to the biologists who make the real experiments on the necessity of knowing some coefficients with a very good accuracy whereas some others can be only roughly known.
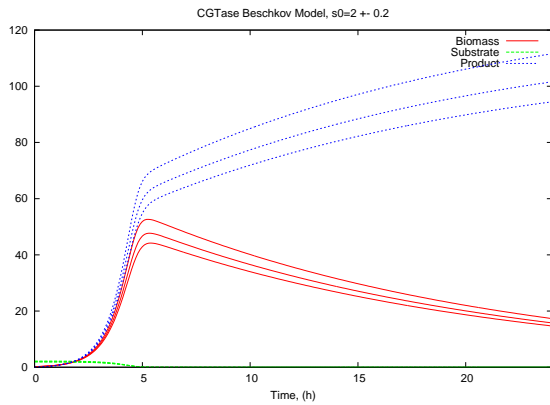
V. FIGURES

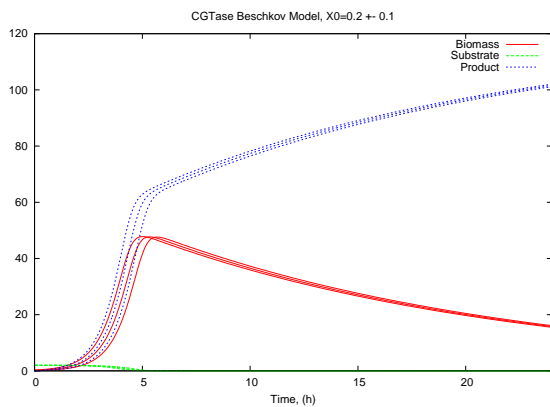Fig. 1: Sensitivity of system (6) to $s_0 = 2 \pm 0.2$

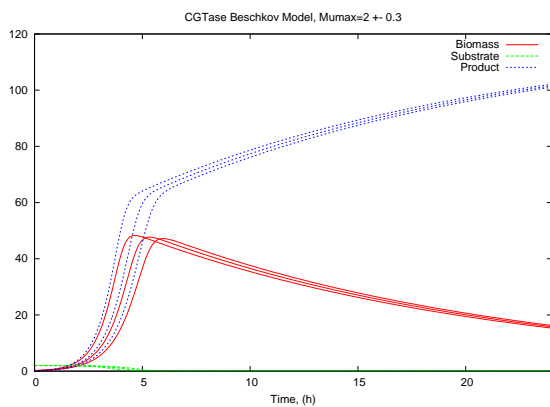Fig. 2: Sensitivity of system (6) to $x_0 = 0.2 \pm 0.1$

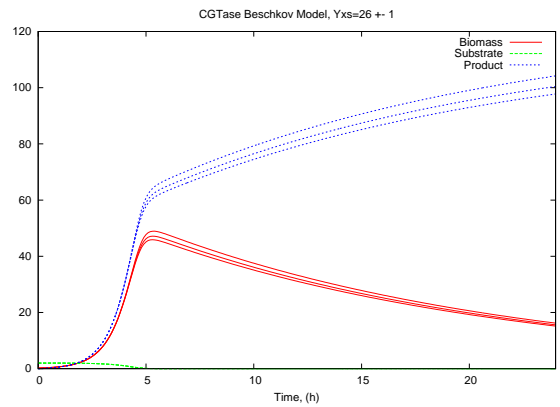Fig. 3: Sensitivity of system (6) to $\mu_{max} = 2 \pm 0.3$

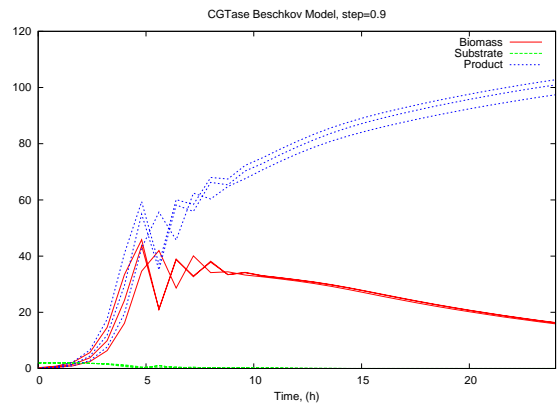Fig. 4: Sensitivity of system (6) to $Y_{X/S} = 26 \pm 1$

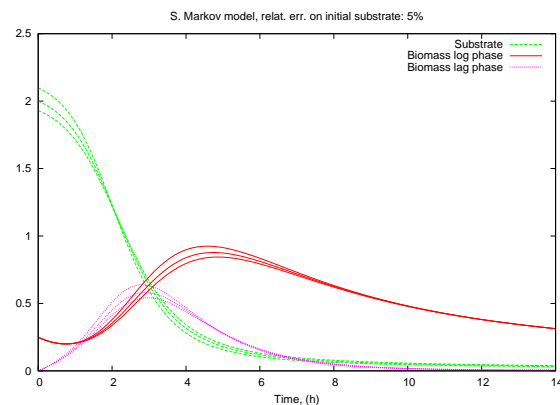Fig. 5: Solution of system (6) with integrating step $h = 0.9$
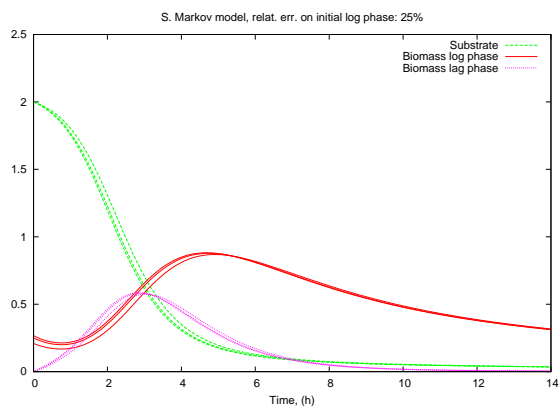

Fig. 6: Sensitivity of system (7) to $s_0 = 2 \pm 5\%$
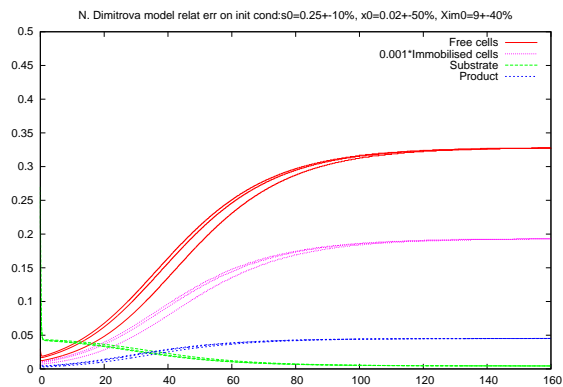
Fig. 7: Sensitivity of system (7) to $x_0 = 0.25 \pm 25\%$

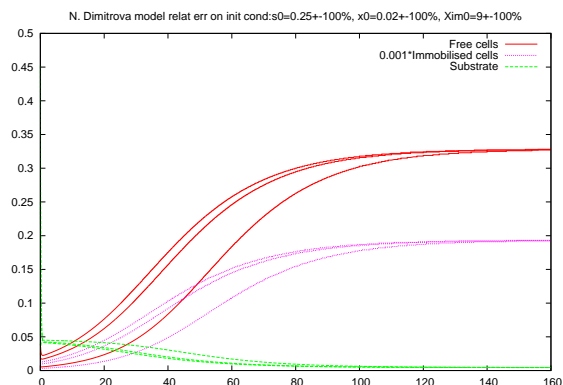Fig. 8: Sensitivity of (8) to $x_1(0)$, $x_{im}(0)$, $s(0)$

Fig. 9: Sensitivity of (8) to 100% errors on initial conditions

REFERENCES

[1] R. Alt, S. Markov, Theoretical and computational studies of some bioreactor models, *Computers and Mathematics with Applications* 64 (2012), 350–360.

[2] R. Alt, J.-L. Lamotte, S. Markov, Stochastic Arithmetic, theory and experiments, *Serdica J. Computing 4* (2010), 101–110.

[3] R. Alt, J. Vignes, Stochastic arithmetic as a model of granular computing. In W. Pedrycz, A. Skowron, and V Kreinovitch, editors, Handbook of Granular Computing, chapter 2. Wiley and sons, 2008. http://dx.doi.org/10.1002/9780470724163

[4] R. Alt, J.-L. Lamotte, S. Markov, Numerical study of algebraic problems using stochastic arithmetic, In I. Lirkov, S. Margenov, J. Wasniewski (Eds.), Large-Scale Scientific Computing, LNCS 4818, Springer (2008), 123–130. http://dx.doi.org/10.1007/978-3-540-78827-0_12

[5] R. Alt, J.-L. Lamotte, S. Markov, Abstract structures in stochastic arithmetic, In B. Bouchon-Meunier, R. R. Yager (Eds.), Proc. 11-th Conference on Information Processing and Management of Uncertainties in Knowledge-based Systems (IPMU'06), Edit. EDK, Paris, 2006, 794–801.

[6] M. Borisov, N. Dimitrova, V. Beschkov, Stability Analysis of a Biorector Model for Biodegradation of Xenobiotics.*Computers and Mathematics with Applications*, vol. 64, No. 3, 2012, 361–373. http://dx.doi.org/10.1016/j.camwa.2012.02.067

[7] N. Burhan, Ts. Sapundzhiev, V. Beschkov, Mathematical modelling of cyclodextrin-glucanotransferase production by batch cultivation. *Biochemical Engineering Journal 24*, 2005, 73–77

[8] http://www-pequan.lip6.fr/cadna//

[9] J.M. Chesneaux, Study of the computing accuracy by using a probabilistic approach, *Contribution to Comp. Arith.and Self validating methods*, C.Ullrich Ed, IMACS, N.J.,1990, 19–30.

[10] M. La Porte, J. Vignes, Etude statistique des erreurs dans l'arithmétique des ordinateurs, application au contrôle des résultats d'algorithmes numériques, *Numer. Math.*, 23, 1974, 63–72

[11] S. Markov, R. Alt, J.L. Lamotte, Stochastic arithmetic: S-spaces and some applications, *Numer. Algo.* 37 (1–4), 2004, 275–284. http://dx.doi.org/10.1023/B:NUMA.0000049474.51465.41

[12] L. Markus, H.Y. Yamabe, Globals stability criteria for differential systems, *Osaka Math. J.* 12, 1960, 305–317.

[13] S. Markov, R. Alt, Stochastic arithmetic, addition and multiplication by scalars, *Appl. Numer. Math*, 50, 2004, 475–488.

[14] J. Vignes, A stochastic arithmetic for reliable scientific computation, *Math. and Comp. in Sim.* 35, 1993, 233–261.