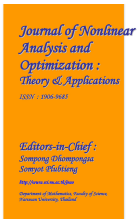
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 2, No. 1, (2011), 119-129
ISSN : 1906-9605
http://www.sci.nu.ac.th/jnao



UNCONSTRAINED OPTIMIZATION IN A STOCHASTIC CELLULAR AUTOMATA SYSTEM

H. C. JIMBO[1,*], M. J. CRAVEN[2]

[1]NARA Institute of Science and Technology, Ikoma, 8916 - 5 Takayama, 630-0192 Nara, Japan

[2]School of Computing and Engineering Systems, University of Abertay, Bell Street, Dundee DD1 1HG, UK

ABSTRACT. This paper considers a stochastic cellular automata system which models a random dynamical system, and introduces a simple unconstrained optimization problem on such a system to capture hidden characteristics over time. To achieve this goal, we create a random metric which is applied to nearby and faraway locations of automata in order to find hidden characteristics in the automata system over time. Solving the random metric based unconstrained optimization problem, we found that solutions show high and low level fluctuations, depending on the choice of the perturbation parameter λ and the corresponding locations. The application of our method to cell concentration data reveals its consistency and adaptability. This work is an expanded version of our previous work [5].

KEYWORDS : Unconstrained optimization; Nonlinear dynamic; Time series analysis; Local autoregressive modeling; Probabilistic metric.

1. INTRODUCTION

In the last two decades, the analysis of various properties of dynamical systems has been a field of active research. Researchers have now agreed that dynamical models may be analyzed by estimating the frequency spectrum of the system. This may be done by non-parametric methods; for example, Fast Fourier Transform [10] or by fitting the autoregressive model, which is a parametric model [7]. However, in the presence of a high-dimensional system with hidden characteristics, the first

*Corresponding author.
Email address : jimbohc@bs.naist.jp (H. C. Jimbo).
Article history : Received 15 December 2010. Accepted 1 February 2011.

approach is inappropriate because it does not explain the temporal events. In contrast, the second approach provides a more flexible framework for parsimonious dynamical modeling of time series data, which can be readily applied to prediction or classification. There are many dynamical phenomena in nature that may be treated as stochastic processes; for example financial markets, cell dynamics, hydrodynamics, economic, biological and psychological events. Although most of them may not be considered as stochastic at the microscopic level, they are to be considered as such at macroscopic level. By sampling the dynamical process at equally spaced intervals of time we characterize the stochastic processes and their predictive mechanisms [3, 6].

Without loss of generality, we consider a dynamical system with n particles randomly interacting over time, with on and off states respectively. Here, on and off states depend on the type of system being used: "on" may be an activated state whereas "off" may be a non-activated state. If we consider a biological process, for example, then "on" may be a high concentration and "off" a low concentration of some chemical or substance. We are interested in finding the optimal method of extracting hidden information among those particles, so as to be able to understand and predict the future behavior of the system by optimization of such information. As a first step we should be able to capture the nonlinearity in the dynamic and measurability of the random process which is involved. But, nonlinearity itself is not a property, but rather the absence of property [10], and therefore we have to be more specific as to our interpretation of nonlinearity. We adopt the notation $X_i = X_{i,t} = X_i(t)$, read as "the state X at location i at time t". Consequently we could build a class of parametric models based on the reconstructed state space (for details, see [11]), but here we prefer to explore the direct parametric model of the state dynamics by using the time series $X_i(t)$, $i = 1, \ldots, N$ at various "ON" and "OFF" locations (analogously to [5]). Further, we consider as a modeling assumption that each particle in the system *de facto* interacts with "nearby" and/or "faraway" neighbors where interactions are captured by our designed metric. Furthermore, we define the state dynamic as a mixture of the dynamics at the i-th locations where the "on state of the system" is represented by X_i and those at the j-th locations represented by X_j (or the "off state of the system"). We shall explore the linear dependence among local states using a linear function f of the past states of the stochastic cellular automata system; this leads us to the stochastic equations

$$X_i = f\left(X_{i-1}, \ldots, X_{i-p}\right) + \varepsilon_i \tag{1.1}$$

$$X_j = f\left(X_{j-1}, \ldots, X_{j-q}\right) + \varepsilon_j \tag{1.2}$$

Above, p and q are the respective model orders and ε_i and ε_j represent the respective dynamical noises of the system. If $f(\cdot)$ is a linear function, we obtain the classic $AR(p)$ models. The correlation $\rho\left(X_i, X_j\right)$ between random states X_i and X_j is statistically given by

$$\rho\left(X_i, X_j\right) = \rho(i,j) = \begin{cases} \rho_{ij} & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}, \tag{1.3}$$

where

$$\rho(i,j) = \frac{E\left(x_i - Ex_i\right)\left(x_j - Ex_j\right)}{E\left(x_i - Ex_i\right)^2 E\left(x_j - Ex_j\right)^2} \tag{1.4}$$

and $i \cong [\Delta t]$ and $j \cong [\Delta t]$. The last approximation occurs because, as in [1, 4, 11], we may obtain a stationary model from any stationary nonlinear stochastic dynamical system model (as sampling intervals are fixed in time series models). In the next section we will explicitly define the function f which is important in designing our quantitative measure.

2. LOCAL AUTOREGRESSIVE MODELS

In order to design a quantitative measure for such a dynamical system, we assume that nonlinearity may be modeled by a nonlinear as well as a linear model with some added noise (as in [10]). Here we choose the following Local Autoregressive Models (LAMs) of orders p and q respectively (for on/off locations):

$$X_i = \sum_{k=1}^{p} a_k X_{i-k} + \varepsilon_i \qquad i - AR(p) \tag{2.1}$$

$$X_j = \sum_{l=1}^{q} b_l X_{j-l} + \varepsilon_j \qquad j - AR(q) \tag{2.2}$$

where $\varepsilon_i \sim N(0,1), \varepsilon_j \sim N(0,1)$ are Gaussian noises of the system. We define a bandwidth parameter h such that $X_i = \frac{\alpha X_{j-1}}{h}$ approaches zero for the largest occurring values of h. Observe that h is a variable parameter and α is a fixed parameter. In the next section we state three lemmas to support our method.

Lemma 2.1 (Kolmogorov-Smirnov)**.** *The distance between two independent random variables is the distance between their probability distribution functions:*

$$d(X_i, X_j) = \sup_{x_i, x_j} |P(X_i \leq x_i) - P(X_j \leq x_j)| \tag{2.3}$$

This lemma provides a useful way of checking equality of certain functions of independent identically distributed random variables, where x_i, x_j are values of random states X_i, X_j, and $P(X_i \leq x_i)$ and $P(X_j \leq x_j)$ are the respective probabilities of those values.

Lemma 2.2. *The distance between two random states X_i and X_j is equal to the p-th mean of the consecutive difference between the probable values of those random variables:*

$$d_p(X_i, X_j) = E_P |X_i - X_j|^p = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{N} (x_i p_i - x_j p_j)^p \tag{2.4}$$

This is known as the *mean difference value* in probability theory. The proofs to Lemmas 2.1 and 2.2 may be found in [8].

Lemma 2.3. *The distance between two independent and identically distributed random variables X_i and X_j, with given means and standard deviations is equal to the expected value of their consecutive differences.*

$$d_2(X_i, X_j) = E_P |X_i - X_j|^2; \qquad |i - j| \leq 1 \tag{2.5}$$

The proof of Lemma 2.3 follows from our results in Sections 3 and 4.

In the case when there are some perturbations in the system (that is, there exists at least one pair (i,j) such that $\lambda_{ij} \neq 0$) and the states X_i and X_j are faraway from each other i.e. $|i-j|>1$ we write:

$$d_2\left(X_i, X_j\right) = E_{P_0}\left|X_i - X_j\right|^2 + \sum_{i=1}^{N}\sum_{j=1}^{N} \lambda_{ij}\left|P(\omega_i) - \lambda_{ij}P'(\omega_j)\right|; |i-j|>1, i\leq N, j\leq N \tag{2.6}$$

where $P_0 = \left(P_i, P_j'\right)$ is a simple scalar product, P represents a probability at location i and P' represents a probability at location j. Observe that Lemma 2.3 is a consequence of Lemmas 2.1 and 2.2, and it will be used in the next section of the paper.

3. MEASURABILITY

3.1. **Mathematical Foundation.** After identifying the system (see [5]), we build a metric to measure the degree of interactions between the two state dynamics X_i and X_j. We develop an analogue of Kolmogorov distance over the sample distributions at locations i and j, called an α-*metric*. Technically, we assume that minimizing such a metric will provide a solution which corresponds to the hidden characteristics of the system; this is justified by the law of minimal entropy in physics. It is clear that nearby states (locations such that $|i-j|\leq 1$) will be less perturbed than faraway states (locations such that $|i-j|>1$); thus our α-metric is

$$J(i,j,\lambda) = \begin{cases} E_{P_0}|X_i - X_j|^2; & |i-j|\leq 1 \quad \text{(i)} \\ E_{P_0}|X_i - X_j|^2 + \sum\limits_{l=1}^{N} \lambda_{kl}\left|P(\omega_l) - \lambda_{kl}P'(\omega_l)\right|; & |i-j|>1 \quad \text{(ii)} \end{cases} \tag{3.1}$$

where

$$\begin{aligned} E_{P_0}\left|X_i - X_j\right|^2 &= \frac{1}{N-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|X_i - X_j\right|^2 P_0\left(X_i < x_i, X_j < x_j\right) \\ &= \left|x_{i1}p_1 - x_{j1}p_1'\right|^2 + \left|x_{i2}p_2 - x_{j2}p_2'\right|^2 + \ldots + \left|x_{in}p_n - x_{jn}p_n'\right|^2 \end{aligned}$$

and $X_i = X(\omega_i) = x_i$ with probability $P(\omega_i) = p_i \in (0,1)$, $X_j = X(\omega_j) = x_j$ with probability $P(\omega_j) = p_j \in (0,1)$ with

$$E_P X_i = \sum_{k=1}^{n} x_i p_k = \sum_{k=1}^{n} x_{ik}p_k = x_{i1}p_1 + x_{i2}p_2 + \ldots + x_{in}p_n \tag{3.2}$$

$$E_{P'} X_j = \sum_{l=1}^{n} x_j p_l = \sum_{l=1}^{n} x_{jl}p_l = x_{j1}p_1' + x_{j2}p_2' + \ldots + x_{jn}p_n'. \tag{3.3}$$

From (3.1), subequation (i) is the analogue of Kolmogorov distance (KD), and subequation (ii) is an adjusted version of KD called (AKD) and λ is the parameter of regularization of the interactions between particles (see [2]). We will discuss the following cases:

(a) If $\lambda = 0$, then (i) and (ii) are identical and $|i-j|\leq 1$. That is, J represents the probabilistic metric of nearest neighboring states.

(b) If $\lambda \neq 0$ then $J_1 = E_P |X_i - X_j|^2 + \lambda E_P |X_i - X_j|^2$ is the probabilistic metric of the faraway neighboring states. Finally, we wish to find the solution to the minimization problem

$$\widetilde{J} = \min_{x_i, x_j, \lambda} J(x_i, x_j, \lambda). \tag{3.4}$$

At each given time, (14) gives an unconstrained linear optimization problem on the random states X_i and X_j, which can be easily solved.

Proposition 3.1. *If there are perturbations in the system, the dynamic quadratic variation of both X_i and X_j is proportional to their probabilistic variation in the perturbed part of the system: $|X_i - X_j|^2 = \lambda |P_i - P_j|$.*

The above proposition is important because it simplifies the manipulation of the automata system at the perturbed locations. In this way, we may create a simulation using approximations of various equations, which we shall observe in the next subsection.

3.2. **Solving the Unconstrained Optimization Problem.** We wish to solve the equation (3.4), where

$$J = \frac{1}{N-1} \sum_{i=1}^{N} \sum_{j=1}^{N} |X_i - X_j|^2 P(X_i < x_i, X_j < x_j) + \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_j |P(\omega_i) - \lambda_j P(\omega_j)|, \tag{3.5}$$

which we decompose in two parts in an intuitive way: $J = J_1 + J_2$. The first part, J_1, of this equation gives the approximation

$$\begin{aligned} & \sum_{i=1}^{N} \sum_{j=1}^{N} |X_i - X_j|^2 P(X_i < x_j, X_j < x_j) \\ \approx \quad & (x_{i1} p_1 - X_{j1} p_1')^2 + (x_{i2} p_2 - X_{j2} p_2')^2 + \ldots + (x_{in} p_n - X_{jn} p_n')^2 \end{aligned} \tag{3.6}$$

and the second part may be approximated as

$$\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_j |P(\omega_i) - \lambda_j P(\omega_j)| \approx \lambda_1 (p_1 - \lambda_1 p_1')^2 + \lambda_2 (p_2 - \lambda_2 p_2')^2 + \ldots + \lambda_n (p_n - \lambda_n p_n')^2. \tag{3.7}$$

Taking the first partial derivative of J with respect to the i-th location, and putting the sum equal to zero implies that the optimal x_i is $\hat{x_i} = \left(\frac{p_1}{p_1'}, \frac{p_2}{p_2'}, \ldots, \frac{p_n}{p_n'} \right)$. Analogously, with respect to the j-th location we have $\hat{x_j} = \left(\frac{p_1'}{p_1}, \frac{p_2'}{p_2}, \ldots, \frac{p_n'}{p_n} \right)$. Partially differentiating all the components of J with respect to λ_m gives $\hat{\lambda} = \left(\frac{p_1}{2p_1'}, \frac{p_2}{2p_2'}, \ldots, \frac{p_n}{2p_n'} \right)$. Finally, to check whether our optimal solutions $\left(\hat{x_i}, \hat{x_j}, \hat{\lambda} \right)$ are consistent and can be validated in general, we apply our new concept to existing cell concentration data to obtain the results presented in the next section.

4. SIMULATION AND RESULTS

To validate our theoretical framework, we apply our concept to real existing biological data. We use the cellular concentration data available at [13] and apply our method to measure the optimal adjusted Kolmogorov distance between states at various locations over time. We let the initial states be $X_i(0)$, the initial cell concentration at the i-th locations (which we take to be the first column of the

aforementioned data), and $X_j(0)$, the initial cell concentration at the j-th locations (we take this to be the second column of the data). The "on" and "off" states are respectively represented by increases and decreases of the concentration at a given location over time. First we plot the dynamic at each location and the α-metric in separate graphs.

We subdivide our cases studied into five classes based on the choice of the value of lambda, which is defined in this work as the perturbation parameter and is applied only to faraway states. All dynamics for cases (a) and (b) show fluctuations at the beginning but those fluctuations become rapidly damped over time. We now plot each scenario based on the choice of $\lambda_{ij} \in [0, 1]$. The results in this section are taken from [5].

4.1. **Extremal Cases $\lambda_{ij} = 0$ or $\lambda_{ij} = 1$.** We take λ_{ij} to be either of the endpoints of the closed interval $[0, 1]$. Figs. 1 and 2 give typical plots for the dynamic of X_i, the dynamic of X_j and the dynamic of the α-metric over time, respectively.

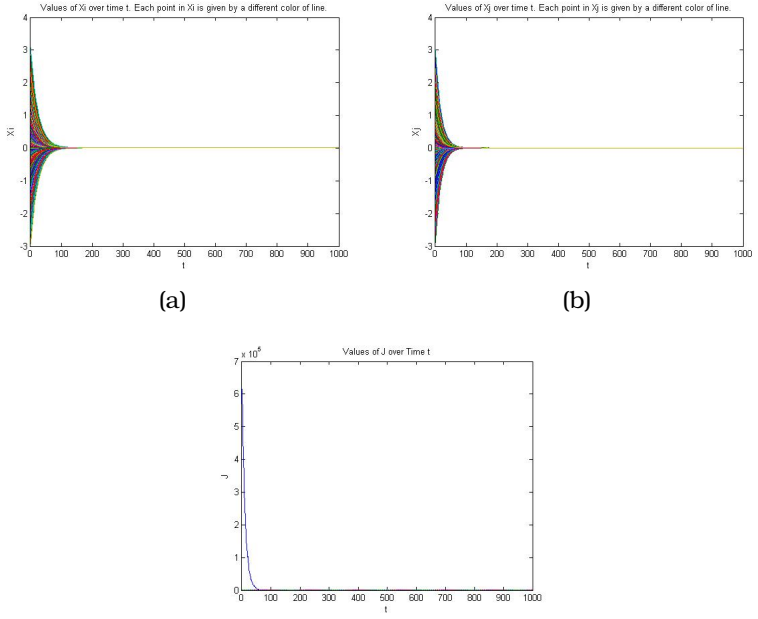


FIGURE 1. Case 1: $\lambda_{ij} = 0$ for all i, j.

4.2. **Normal Cases $\lambda_{ij} \neq 0, 1$.** This time we take λ_{ij} to be to be one of three values not equal to the endpoints of the interval. Figs. 3, 4 and 5 give typical plots for the dynamic of X_i, the dynamic of X_j and the dynamic of the α-metric over time, respectively.

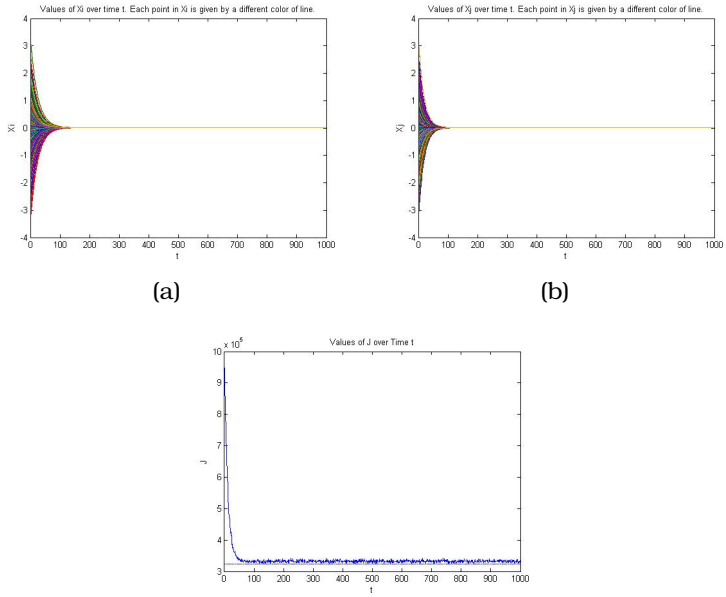


FIGURE 2. Case 2: $\lambda_{ij} = 1$ for all i, j.

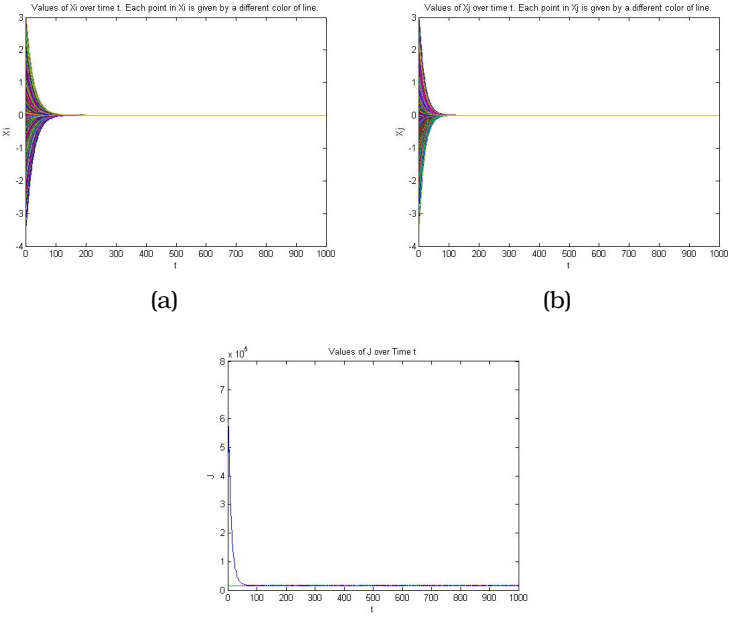
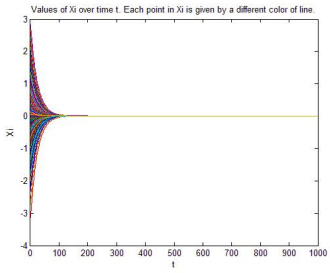
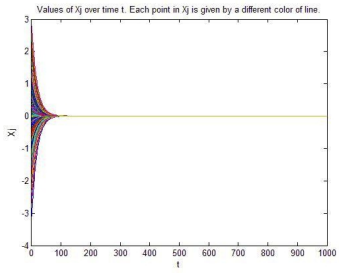


FIGURE 3. Case 3: $\lambda_{ij} = \lambda_N = N-$dimensional constant vector.

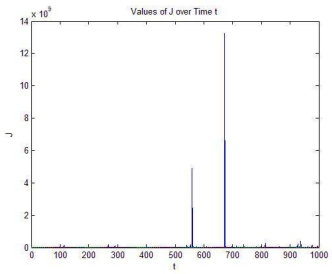
Comments: Figs. 1–5 are detailed plots of the dynamics of particles $X_i(t)$ and $X_j(t)$ over time and the α-metric J over 1000 timesteps. Observe that cases 1, 2



(a)

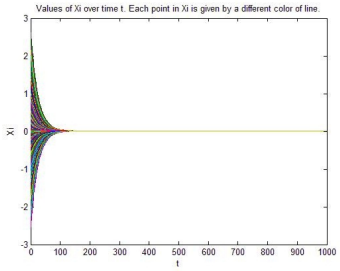


(b)

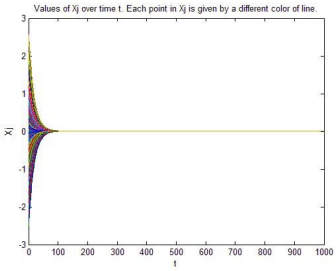


(c)

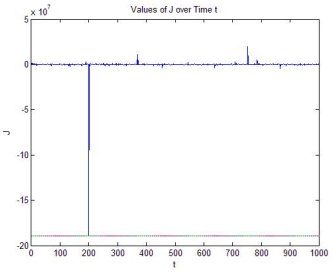
FIGURE 4. Case 4: $\lambda_{ij} = \lambda_i$.



(a)



(b)



(c)

FIGURE 5. Case 5: $\lambda_{ij} = \lambda_j$.

and 3 exhibit low fluctuations in the α-metric J over time, but in the remaining cases there are very high fluctuations in J. However it is surprising to observe

that the smallest value of J is achieved faster in case 5. This again proves that random fluctuations may be beneficial to dynamical systems in terms of optimality and stability. Additionally, this example shows that local stability drives global stability.

In the following tables, we give the optimal value $\widetilde{J}$ and the standard deviation of the α-metric J found in each case.

Extremal Cases	1. $\lambda_{ij} = 0$	2. $\lambda_{ij} = 1$
$\overline{J}$	4.765×10^4	3.226×10^5
SD	1.637×10^1	4.667×10^4

TABLE 1. This table shows the minimum value of the adjusted Kolmogorov distance for each perturbation λ_{ij} in the extremal cases. Comparing both cases, we observe that case 1 has the smallest optimal J and spread (or SD). This means that although case 1 has the best optimal J, the number of states reaching such optimality are very limited. In case 2, J attains an optimal value greater than that of case 1, but with a greater spread. These results confirm relative stability of the cell concentration stochastic cellular system, since optimal J exist in both cases and are comparable.

Normal Cases	3. $\lambda_{ij} = \lambda_N$	4. $\lambda_{ij} = \lambda_i$	5. $\lambda_{ij} = \lambda_j$
$\overline{J}$	1.565×10^4	5.232×10^5	-1.893×10^8
SD	4.960×10^4	4.465×10^8	6.047×10^6

TABLE 2. This table shows the minimum values of the adjusted Kolmogorov distance for each perturbation λ_{ij} in the regular cases together with the spread. Case 5 has the smallest optimal j (the fact it is so small is interesting in itself) with a medium spread, and case 4 has the medium optimal J but with very high spread. Finally case 3 has a low optimal J (comparable to case 1) and also the smallest spread among the normal cases. These results indicate that many states in case 3 reach the optimal level, and that states at the j-th locations are more robust than those at the i-th locations.

Further, if we compare the normal and extremal cases we observe that the spreads for the normal cases are at least those of the extremal cases, but comparing optimal values in both normal and extremal cases remains difficult in general (optimal J in case 1 is relatively stable, but is unstable in case 2).

5. DISCUSSION AND CONCLUSION

In this paper, we have designed a random metric (the so-called α-metric) to capture nonlinearity and important hidden events in a "random" dynamical system. To this end we have incorporated a Local Autoregressive Model of the dynamic into the "probabilistic" metric and solved an unconstrained optimization problem on a stochastic cellular automata system. We hypothesize the solution will correspond to some hidden characteristics of the state dynamics. Such information may prove to be useful for exploring such a system. Again our motivation here comes from the wish to optimally capture hidden but relevant information in a stochastic cellular automata system defined as a mixture of stochastic processes.

Further, the created random metric was applied to nearby and faraway locations in order to find hidden characteristics of the automata system over time. Solving the unconstrained optimization random metric based problem, we found that solutions show high and low level fluctuations depending on the choice of the perturbation parameter λ and the location. The application of our method to cell concentration data reveals its consistency and adaptability (Table 2). Finally this exercise confirms that reliable detection and quantitative description in stochastic systems with limited precision remains a difficult task, especially when the dynamic is characterized by some additional complexity. But the example here with cellular data shows appreciable results despite a rather high dimension; therefore our methodology which combines a random metric and unconstrained optimization has the merit of producing interesting results.

Acknowledgements. H. C. Jimbo gratefully acknowledges the generous support given by the GCOE International Research Fellowship. We thank NAIST (Nara Institute of Science and Technology, Nara, Japan) and the University of Abertay for their generous support during the preparation of this work. We also thank the anonymous referees for the helpful comments from which some sections of this work were improved.

REFERENCES

1. M. Barahona, C. S. Poon, Detecting of Nonlinear Dynamic in Short Noise Time Series, Nature 381 (1996), 215-217.
2. C. M. Bishop, Pattern Recognition and Machine Learning, Information Science and Statistics, Springer (2006).
3. W. Feller, An Introduction to Probability theory and its Applications (Vol. 2), John Wiley and Sons (1972).
4. R. Hegger, H. Kanty, T. Schreiber, Practical implementation of nonlinear time series methods: The TISEAN package, CHAOS 9(1999)
5. H.C, Jimbo, M. J. Craven, Minimizing Kolmogorov Distance in Stochastic Cellular Automata Models, Proc. Neural, Parallel and Scientific Computations 4 (2010), 187-192.
6. H. C. Jimbo, Distribution Characterization in a Practical Moment Problem, Acta Mathematica 73, no. 1 (2004), 107-114.
7. S. M. Kay, Modern Spectral Estimation: Theory and Application, Prentice-Hall, Englewood Cliffs (1988).
8. A. N. Kolmogorov, Selected Works of A. N. Kolmogorov (Vol. II): Probability Theory and Mathematical Statistics (Translation by G. Lindquist of the 1986 original), Kluwer, Dordrecht (1992).

9. M. Levena, J. Bar-Ilan, Computer Journal, UK (50) 567 (2007).
10. M. B. Priestley, Spectral Analysis and Time Series Analysis (Vols. I and II), Academic Press (1981).
11. J. Pearl, Probabilistic Reasoning in Intelligence Systems: Network of plausible Inference, Morgan Kaufmann (1989).
12. D. Sonnette, Critical Phenomena in Natural Science (2nd. Ed.), Springer Verlag (2003).
13. NAIST Website http://bsw3.naist.jp/tns/labMembers/ResearchFellow/relative Cell Concentrations.pdf (2010).