# Multifractal modeling of the production of concentrated sugar syrup crystal

Sheng Bi(闭胜) and Jianbo Gao(高剑波)†

*Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning* 530005, *China*

(Received 12 January 2016; revised manuscript received 29 February 2016; published online 31 May 2016)

High quality, concentrated sugar syrup crystal is produced in a critical step in cane sugar production: the clarification process. It is characterized by two variables: the color of the produced sugar and its clarity degree. We show that the temporal variations of these variables follow power-law distributions and can be well modeled by multiplicative cascade multifractal processes. These interesting properties suggest that the degradation in color and clarity degree has a system-wide cause. In particular, the cascade multifractal model suggests that the degradation in color and clarity degree can be equivalently accounted for by the initial "impurities" in the sugarcane. Hence, more effective cleaning of the sugarcane before the clarification stage may lead to substantial improvement in the effect of clarification.

**Keywords:** time series analysis, non-poisson process, power-law distribution, multiplicative cascade multifractal processes

**PACS:** 05.10.Gg, 05.30.Pr, 05.40.Ca, 05.45.Df **DOI:** 10.1088/1674-1056/25/7/070502

## 1. Introduction

Cane sugar (sucrose) is a basic livelihood commodity produced from sugarcane. Among the major cane sugar producers and consumers, China ranks the 3rd, accounting for 7% of the world total. Within China, the Guangxi Autonomous Region in southwest China is the biggest producer of cane sugar in the country, accounting for more than 60% of the market. Unlike in developed countries such as USA and Australia, cane sugar production in most factories in Guangxi and many other regions in China has not been made automatic, partly due to a lack of understanding of the fundamental physics and chemistry involved in the production process.

In general, cane sugar processing consists of six phases: milling, clarification, evaporation, crystallization, centrifuging, and drying so as to obtain white sugar, brown sugar, and other products.[1] In particular, the juice from the milling workshop is called the mixed juice, which contains water, sucrose, bagasse, soil, sand, reduced sugar, and many organic and inorganic non-sugar components. The latter are residual nutrients in the sugarcane, including colloidal substances, inorganic salts (iron, magnesium, aluminum, calcium, and so on), and pigments. Affecting the appearance, color, and concentration of the sugar, the non-sugar components are detrimental to the sugar production, and thus should be carefully removed. This is the aim of the next stage, the clarification stage. It is characterized by two target variables: the color of the produced sugar and its clarity degree. A typical example of the temporal variation of these variables is shown in Fig. 1, where the time covered is the entire harvest season of sugarcane in Gaungxi, China, which is slightly less than 4 months. The long-range correlations in these time series are thought to hinder the automation of cane sugar production from sugarcanes, as long-range correlations mean that persistent deviations from the target values of these variables may last for a considerable period of time.[1] The quality of clarification is characterized by the smallness in color value and the largeness in clarity degree.

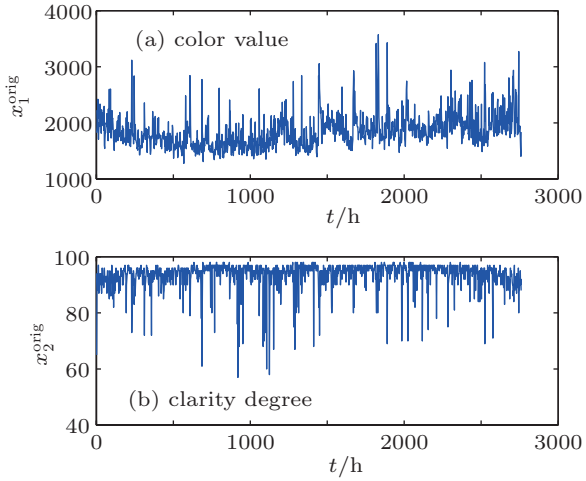



**Fig. 1.** (color online) Raw data of (a) color value $x_1^{\rm orig}$ and (b) clarity degree $x_2^{\rm orig}$.

There has been a great deal of effort expended to improve the clarification of the mixed juice by using physical and chemical means, including using enzymes to reduce viscosity,[2] improving the flocculants,[3] employing electrocoagulation,[4] and enhancing decolorization by hydrodynamic cavitation.[5] These approaches are hoped to directly or indirectly enhance the absorption of impurities by flocs and sediments in the settler. Unfortunately, the effectiveness of the

†Corresponding author. E-mail: jbgao.pmb@gmail.com

© 2016 Chinese Physical Society and IOP Publishing Ltd 

methods has not been thoroughly validated. As a result, these methods have not been adopted in practice yet.

There also has been effort to use neural network based prediction schemes[6,7] to improve cane sugar production. Being black or gray box-based approaches, however, they have not yielded much understanding of the basic physics and chemistry involved in the clarification process, and thus have not helped much with cane sugar production. Observing Fig. 1, we realize that there are two different ways to improve clarification. One is to drastically upgrade the facility in the cane sugar production factory such that the color value can be made significantly smaller while the clarity degree can be made significantly larger. This approach is costly and not realistic. The other approach is to mainly stick with the available facility, but to manage to make the variations in color and clarity degree as small as possible. In other words, the color value would always be close to the minimal value with the clarity degree close to the maximal value shown in Fig. 1. Albeit not as good as the first scenario, this approach is often more than sufficient, since the observed raw color value and clarity degree already conform to industrial standards. This is the scenario that we will focus on in this work. Specifically, we will use multiplicative cascade multifractal to understand whether the variations in the color value and the clarity degree are purely random, or have system-wide causes. If the latter is the case, then it is very likely that effective means can be devised to substantially improve the clarification of the mixed juice from the milling stage.

The multifractal is one of the most important models contributed by complexity science. It is worth noting with the excitement that there has been a great deal of research on complexity science by the Chinese physical community, including works on stochastic resonance, bifurcations, and chaos.[8–22]

Mathematically, multifractals are characterized by many or infinitely many power-law relations. In this paper, we work with a specific type of multifractal, called the random multiplicative process model, to analyze the color value and the clarity degree. This type of multifractal was initially developed to understand the intermittent features of turbulence.[23–25] Mandelbrot was among the first to introduce this concept. Parisi and Frisch's work[26] on turbulence has made it widely known. It has been applied to the study of various phenomena, such as rainfall,[27] liquid water distributions inside marine stratocumuli,[28,29] finance,[30] tropical deep convective variability,[31] and network traffic.[32,33]

The remainder of the paper is organized as follows. In Section 2, we perform a distribution analysis of the clarification process, to infer whether the variations in the color value and the clarity degree may have system-wide causes. In Section 3, we employ the cascade multiplicative multifractal process to model the variations in the color and clarity degree. We show that those variations can be characterized by multifractal processes. The concluding discussion is presented in Section 4.

## 2. Distribution analysis of the clarification process

As we have pointed out, in order to infer whether the variations in the color value and the clarity degree may have system-wide causes, we may focus on the variations in the color value and clarity degree around the minimal and the maximal values, respectively. Denote the raw data of the color value and clarity degree by $x_1^{\text{orig}}$ and $x_2^{\text{orig}}$. The scenario taken here amounts to studying

$$x_1 = x_1^{\text{orig}} - \min(x_1^{\text{orig}}),$$
$$x_2 = \max(x_2^{\text{orig}}) - x_2^{\text{orig}}.$$

For ease of later discussion, the time series plots of $x_1$ and $x_2$ are shown in Figs. 2(a) and 3(b), respectively.

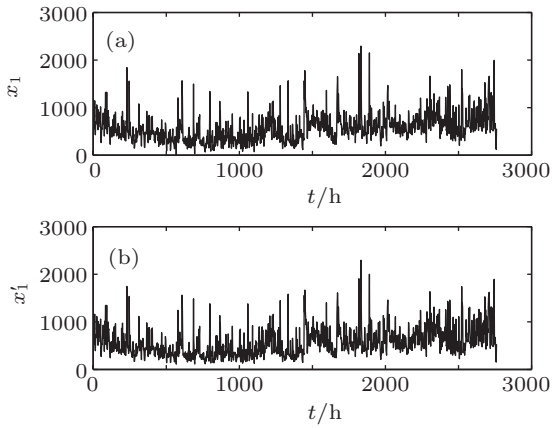



**Fig. 2.** Modeling of the variation of color value: (a) measured and (b) simulated with $\sigma = 0.078$.

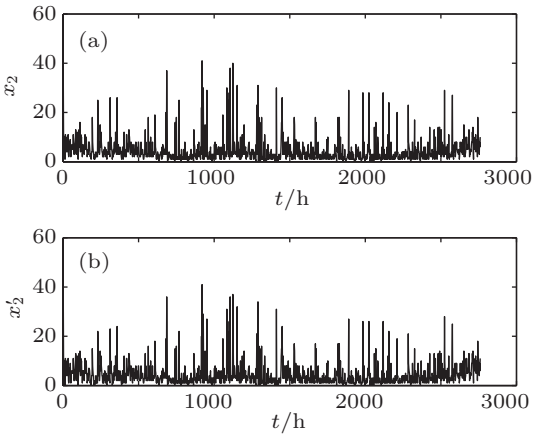



**Fig. 3.** Modeling of the variation of clarity degree: (a) measured and (b) simulated with $\sigma = 0.1385$.

Let us perform a distribution analysis of these variables. Concretely, we will check whether they may follow two specific distributions. One is an exponential distribution, which

may be expressed by its probability density function (PDF)

$$f_X(x) = \lambda\, \mathrm{e}^{-\lambda x}, \quad x \geq 0, \ \ \lambda > 0, \tag{1}$$

or its complementary cumulative distribution function (CCDF)

$$P(X \geq x) = \mathrm{e}^{-\lambda x}, \quad x \geq 0, \ \ \lambda > 0. \tag{2}$$

The other is a power-law distribution, whose PDF is

$$f(x) \sim x^{-\alpha-1}, \quad x \to \infty,$$

and CCDF is

$$P[X \geq x] \sim x^{-\alpha}, \quad x \to \infty. \tag{3}$$

The power-law distribution has a remarkable property that all moments with an order higher than $\alpha$ are infinite. In particular, when $\alpha < 2$, it is called heavy-tailed, since the variance is infinite. When $\alpha \leq 1$, even the mean becomes infinite.

We hypothesize that if the distribution is close to exponential, or other thin-tailed distributions, then the variations in the color value and the clarity degree are simply random and do not have system-wide causes. However, if the distribution is heavy-tailed, then the variations in the color value and the clarity degree may have system-wide causes.

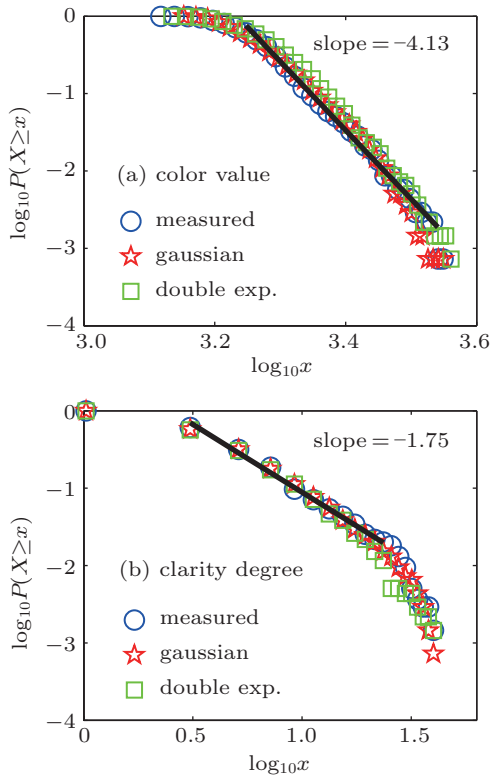



**Fig. 4.** (color online) Complementary cumulative density function (CCDF) for the measured and simulated data with multiplier distributions being Gaussian Eq. (8) and double-exponential Eq. (9). (a) Color value and (b) clarity degree.

We have estimated the distributions for the time series data of $x_1$ and $x_2$. The results are shown in Fig. 4. We observe that both have powerlaw-like distributions. However, the $\alpha$ parameter for the color value is quite large. As a result, the scaling for the powerlaw behavior is valid for a range less than one decade, as shown by the fitted straight line in Fig. 4(a). In contrast, the distribution in the clarity degree is heavy-tailed, with $\alpha < 2$, and a scaling region close to one decade. These analyses suggest that the variations in the color value and the clarity degree may indeed have system-wide causes, and thus may be further analyzed by advanced methods.

## 3. Cascade multifractal analysis and modeling of the clarification process

To gain deeper insights into the clarification of sugarcanes, in this section we perform a multifractal analysis of $x_1$ and $x_2$.

With the framework of the cascade multiplicative model, one considers the moments

$$M_q(\varepsilon) = \sum_i w_i^q, \tag{4}$$

where $q$ is a real number, and the positive "weights" $w_i$ can be readily computed from a time series, as will be explained shortly. One then checks if the following scaling law holds:

$$M_q(\varepsilon) \sim \varepsilon^{\tau(q)}, \ \text{ as } \ \varepsilon \to 0, \tag{5}$$

for different $q$. If the weights $\{w_i\}$ are nonuniform, then the weights $w_i(\varepsilon)$ are said to form a multifractal measure. Note that the normalization $\sum_i w_i = 1$ implies that $\tau(1) = 0$. Also note that if $\{w_i\}$ are uniform, then $\tau(q)$ is linear in $q$. When $\{w_i\}$ are weakly nonuniform, visually $\tau(q)$ may still be approximately linear in $q$. The nonuniformity in $\{w_i\}$ is better characterized by the so-called generalized dimensions $D_q$, defined as

$$D_q = \frac{\tau(q)}{q-1}. \tag{6}$$

Here, $D_q$ is a monotonically decreasing function of $q$.[24] It exhibits a nontrivial dependence on $q$ when the weights $\{w_i\}$ are nonuniform.

To better understand multifractal formalism, we consider below the random multiplicative cascade model. The conservative model, which is most pertinent here, is specified as follows. Consider a unit interval. Associate it with a unit mass. Divide the unit interval into two, say, left and right segments of equal length. Also, partition the associated mass into two fractions, $r$ and $1-r$, and assign them to the left and right segments, respectively. The parameter $r$ is in general a random variable, governed by a PDF $P(r)$, $0 \leq r \leq 1$. The fraction $r$ is called the multiplier. Each new subinterval and its associated weight are further divided into two parts following the same rule. This procedure is schematically shown in Fig. 5, where the multiplier $r$ is written as $r_{ij}$, with $i$ indicating the stage number and $j$ (assuming only odd numbers, leaving even

numbers for $1-r_{ij}$) indicating the positions of a weight on that stage. Note that the scale (i.e., the interval length) associated with stage $i$ is $2^{-i}$. We assume that $P(r)$ is symmetric about $r=1/2$ and has successive moments $\mu_1$, $\mu_2$, .... Hence $r_{ij}$ and $1-r_{ij}$ both have marginal distribution $P(r)$. The weights at the stage $N$, $\{w_n, n=1,\ldots,2^N\}$, can be expressed as

$$w_n = u_1 u_2 \cdots u_N,$$

where $u_l$, $l=1,\ldots,N$, are either $r_{ij}$ or $1-r_{ij}$. Thus, $\{u_i, i\geq 1\}$ are iid random variables having PDF $P(r)$.

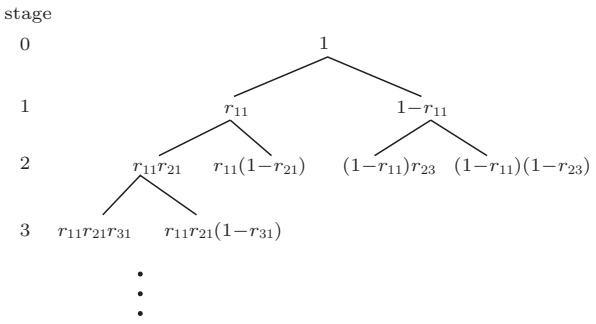



**Fig. 5.** Schematic of the construction rule.

Let us now explain how to obtain the weights $w_i$ from a positive time series. The basic idea in analyzing a generic time series $\{X_i\}$ is to view $\{X_i, i=1,\ldots,2^N\}$ as the weight series of a certain multiplicative process at stage $N$. Under this scenario, the total weight $\sum_{i=1}^{2^N} X_i$ is set equal to 1 unit, and the scale associated with stage $N$ is $\varepsilon = 2^{-N}$. This is the smallest time scale resolvable by the measured data.

Given the weight sequence at stage $N$, the weights at stage $N-1$, $\{X_i^{(2^1)}, i=1,\ldots,2^{N-1}\}$, are obtained by simply adding the consecutive weights at stage $N$ over nonoverlapping blocks of size 2, i.e., $X_i^{(2^1)} = X_{2i-1} + X_{2i}$, for $i=1,\ldots,2^{N-1}$, where the superscript $2^1$ for $X_i^{(2^1)}$ is used to indicate that the block size used for the summation at stage $N-1$ is $2^1$. This follows directly from the construction of a multiplicative multifractal process schematized in Fig. 5. Associated with this stage is the scale $\varepsilon = 2^{-(N-1)}$. This procedure is carried out recursively. That is, given the weights at stage $j+1$, $\{X_i^{(2^{N-j-1})}, i=1,\ldots,2^{j+1}\}$, we obtain the weights at stage $j$, $\{X_i^{(2^{N-j})}, i=1,\ldots,2^j\}$, by adding consecutive weights at stage $j+1$ over nonoverlapping blocks of size 2, i.e.,

$$X_i^{(2^{N-j})} = X_{2i-1}^{(2^{N-j-1})} + X_{2i}^{(2^{N-j-1})}, \tag{7}$$

for $i=1,\ldots,2^j$. Here, the superscript $2^{N-j}$ for $X_i^{(2^{N-j})}$ is used to indicate that the weights at stage $j$ can be equivalently obtained by adding consecutive weights at stage $N$ over nonoverlapping blocks of size $2^{N-j}$. Associated with stage $j$ is the scale $\varepsilon = 2^{-j}$. This procedure stops at stage 0, where we have a single unit weight, $\sum_{i=1}^{2^N} X_i$, and $\varepsilon = 2^0$. The latter is the largest time scale associated with the measured data. Figure 6 shows this procedure schematically.

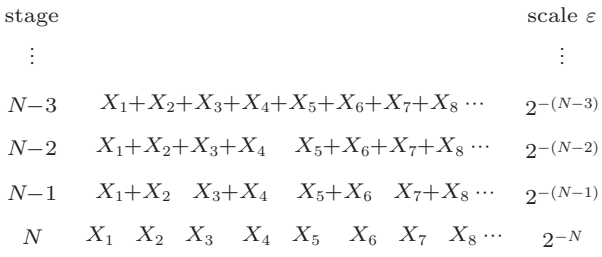



**Fig. 6.** Schematic for obtaining weights from a time series.

Following the above procedure, we have analyzed the multifractal properties of $x_1$ and $x_2$. The results are shown in Figs. 7(a) and 8(a), respectively. We observe that the scaling behaviors are very good. The $\tau(q)$ and $D(q)$ spectra are shown in Figs. 7(b), 7(c) and 8(b), 8(c), respectively. Clearly, both the time series of the color and the clarity degree have multifractal behavior.

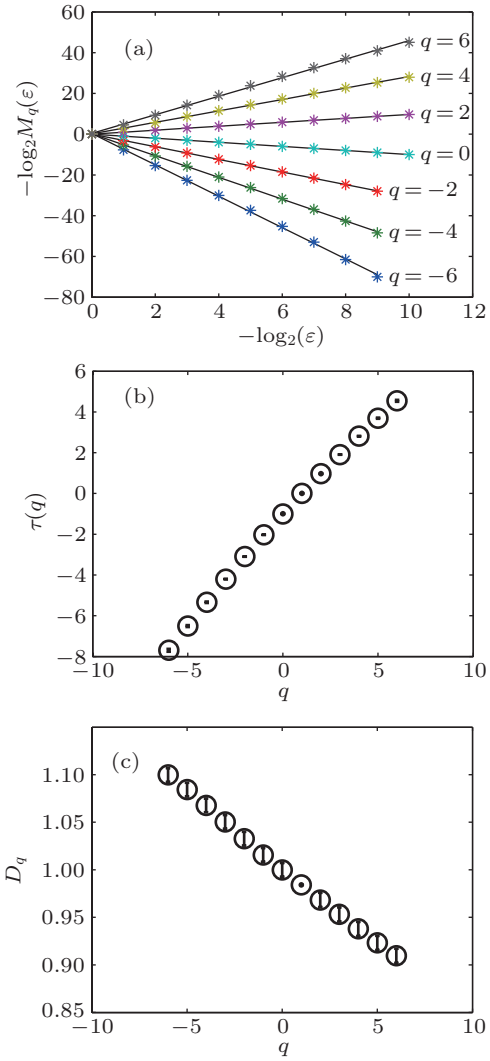



**Fig. 7.** (color online) Multifractal analysis of the time series data of color value. Fitted lines for (a) Eq. (5), (b) $\tau(q)$, and (c) $D_q$.

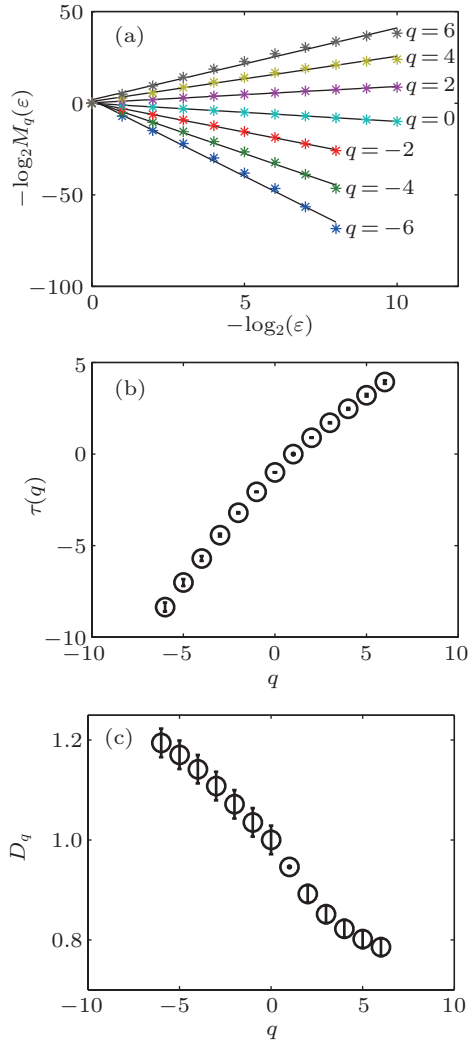



**Fig. 8.** (color online) Multifractal analysis of the time series data of clarity degree. Fitted lines for (a) Eq. (5), (b) $\tau(q)$, and (c) $D_q$

The relevance of the cascade multifractal model to the variations in the color and the clarity degree can be made stronger by actually simulating these variables using the cascade model schematized in Fig. 5. For this purpose, we have used a bell-shaped PDF for the multiplier

$$P(r) = c\,\mathrm{e}^{\frac{-(r-t)^2}{\sigma^2}}, \quad 0 \le r \le 1, \ t = \frac{1}{2}, \tag{8}$$

with mean $t = 1/2$.

The range in the unit interval means that it is a truncated Guassian distribution, therefore, given $\sigma$, $c$ is determined by the condition $\int_0^1 p(r)\mathrm{d}r = 1$. As the time series generated by the cascade model is random, to facilitate a comparison with the measured data, we have re-ordered the generated time series according to the order of the magnitude of each measurement. The results for the simulated color and the clarity degree are shown in Figs. 2(b) and 3(b). We observe that the simulated time series look very similar to the original ones. Indeed, they have very similar distributions, as shown in Fig. 4.

The similarity between the observed and the simulated color and clarity degree data can be further analyzed by examining the long-range correlation property of the data, quantified by the Hurst parameter $0 < H < 1$. It is well-known that depending on whether $H < 1/2, = 1/2$, or $> 1/2$, a time series is said to have anti-persistent correlation, short-range correlation or is memoryless, or long-range correlations. In Ref. [1], we have analyzed the color and the clarity degree data, as well as other variables pertaining to the clarification process using detrended fluctuation analysis (DFA)[34] and adaptive fractal analysis (AFA).[35–38] Here, we only present the results in Fig. 9, leaving the details of DFA and AFA to the reference. Clearly, we observe that the measured and the simulated data have an almost identical long-range correlation property, for time scales up to at least a few days.

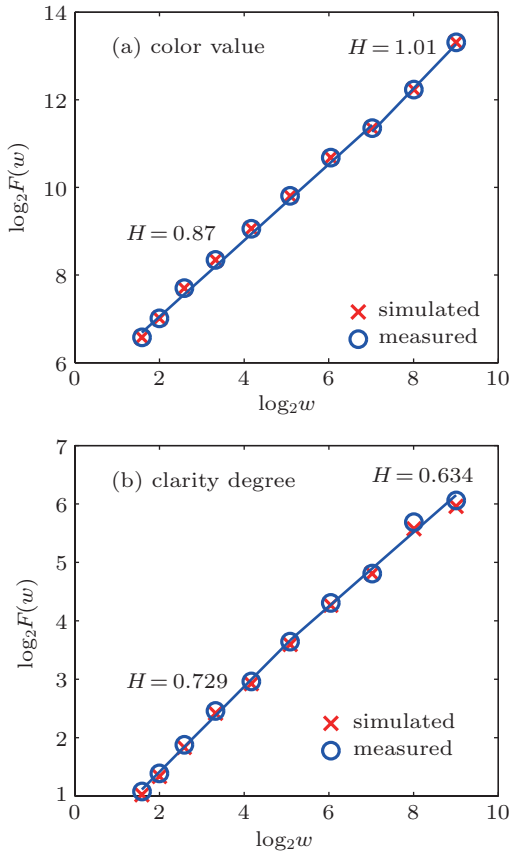



**Fig. 9.** (color online) Long-range correlation analysis of the measured and simulated data. (a) Color value and (b) clarity degree.

In summary, we have shown that the measured color value and clarity degree data possess multifractal properties, and can be well-modeled by the cascade process, in the sense that the distributions and the correlations of the simulated data are almost identical to those of the real data. Interestingly, these features are largely preserved when different functionals are used for the multiplier distribution, including a double-exponential

$$P(r) = c\,\mathrm{e}^{-\alpha_\mathrm{e}|r-1/2|}, \quad 0 \le r \le 1, \tag{9}$$

where $c$ is a constant coeffocient, given by the condition $\int_0^1 P(r)\mathrm{d}r = 1$, when the parameter $\alpha_\mathrm{e}$ is given, a triangular

function,

$$P(r)=\begin{cases}-\dfrac{1}{d^2}r+\dfrac{1}{d}+\dfrac{1}{2d^2}, & 1/2\le r\le 1/2+d,\\ \dfrac{1}{d^2}r+\dfrac{1}{d}-\dfrac{1}{2d^2}, & 1/2-d\le r<1/2,\end{cases} \tag{10}$$

where $0\le d\le 1/2$, and a superposition of uniform distribution and delta function

$$P(r)=\begin{cases}q+p\delta(r-1/2), & 1/2-d\le r\le 1/2+d,\\ 0, & \text{otherwise},\end{cases} \tag{11}$$

where $0\le d\le 1/2, p+2qd=1$. Among these four distributions, the last contains 2 independent parameters, and thus is the most flexible and accurate. These results are consistent with our earlier experience in network traffic modeling.[32,33] To better appreciate these desctiptions, we have compared the measured and the simulated data (Gaussian and double-exponential) in Fig. 4.

## 4. Concluding discussion

Cane sugar production is an important industrial process. One of the most important steps in cane sugar production is the clarification process, which produces high quality, concentrated sugar syrup crystal for further processing. To gain fundamental understanding of the physical and chemical processes associated with the clarification process, and help design better approaches to improve the clarification of the mixed juice, in this paper, we have examined whether the degradation in the color and the clarity degree may have system-wide causes. We have found that the answer is positive. This has further motivated us to employ the cascade multiplicative multifractal model to analyze the variations in the color and the clarity degree. We have shown that those variations can be characterized by multifractal processes. Moreover, they can be conveniently modeled by the cascade model, with the simulated and the observed data having very similar distributions as well as long-range correlation properties.

Broadly speaking, the degradation in the color and the clarity degree may be thought to be caused by the "impurities" in the sugarcane. The cascade model adopted here is essentially a conservative model. This suggests that the "impurities" in the sugarcane may be associated with the "remnants" of the "dirts" of all kinds that remain during the initial cleaning of the sugarcane. Referring to the construction rule of the cascade model shown in Fig. 5, this means that the impurities throughout the entire cane sugar production process can be thought of as coming from the partition of the total impurity (of 1 unit), with the partition rule governed by the multiplier distribution. We surmise that the total impurity is largely introduced by the in-effectiveness in cleaning up the sugarcane in the initial step. If this is so, then to improve clarification, it will be beneficial to use a more powerful cleaning system and filters to better clean up the sugarcane and raw juice. If too much of the impurities are generated by the middle steps during cane sugar production, then the whole cane sugar production process has to be considered fundamentally flawed and in-effective.

## References


[1] Yu X M, Ye J K, Hu J, Liao X P and Gao J B 2013 *Math. Probl. Eng.* **2013** 868313
[2] Li Q T, Huang K N and Pan L L 2010 *Modern Food Sci. Tech.* **26** 619 (in Chinese)
[3] Zhong B J and Gao W H 2010 *Food Sci. Tech.* **9** 022 (in Chinese)
[4] Huang Y C, Guan R C, Yang F and Zhao C Y 2011 *Food Sci. Tech.* **2** 017 (in Chinese)
[5] Huang Y C, Gao H F, Wu X C, Ren X N and Yang F 2014 *Food & Machinery* **30** 5 (in Chinese)
[6] Kamat S, Diwanji V, Smith J G and Madhavan K P 2005 *Proc.* 2005 *IEEE Int. Conf. on Computational Intelligence for Measurement Systems and Applications*, 2005, Giardini Naxos, Italy, pp. 209–214
[7] Lin X F and Yang J R 2009 *IEEE International Joint Conference on Neural Networks*, 2009, Atlanta, Georgia, USA, pp. 1814–1819
[8] Wang Y X, Zhai J Q, Xu W W, Sun G Z and Wu P H 2015 *Chin. Phys. Lett.* **32** 97401
[9] Ji Q B, Yi Z, Yang Z Q and Meng X Y 2015 *Chin. Phys. Lett.* **32** 50501
[10] Zhai J Q, Li Y C, Shi J X, Zhong Y, Li X H and Xu W W 2015 *Chin. Phys. Lett.* **32** 47402
[11] Fang H, Wang Z J, Hong F and Tao G 2015 *Chin. Phys. Lett.* **32** 40502
[12] Tao Y C, Cui M Z, Li H H and Yang J Z 2015 *Chin. Phys. Lett.* **32** 20501
[13] Li G M and Li S X 2015 *Acta Phys. Sin.* **64** 160502 (in Chinese)
[14] Dang X Y, Li H T, Yuan Z S and Hu W 2015 *Acta Phys. Sin.* **64** 160501 (in Chinese)
[15] Wang S T, Wu Z M, Wu J G, Zhou L and Xia G Q 2015 *Acta Phys. Sin.* **64** 154205 (in Chinese)
[16] Ding N, Zhang H J, Zhang X, Ou J W and Luo D 2015 *Acta Phys. Sin.* **64** 139801 (in Chinese)
[17] Wang R and Guo J B 2015 *Acta Phys. Sin.* **64** 130702 (in Chinese)
[18] Ren H P and Bai C 2015 *Chin. Phys. B* **24** 080503
[19] Wei W and Zuo M 2015 *Chin. Phys. B* **24** 080501
[20] Liu S, Wang Z L, Zhao S S, Li H B and Li J X 2015 *Chin. Phys. B* **24** 074501
[21] Shu J 2015 *Chin. Phys. B* **24** 060509
[22] Fang Y, Wang G Y and Wang X Y 2015 *Chin. Phys. B* **24** 060506
[23] Frisch U 1995 *Turbulence—The Legacy of A.N. Kolmogorov* (Cambridge: Cambridge University Press)
[24] Gouyet J F 1995 *Physics and Fractal Structures* (New York: Springer)
[25] Frederiksen R D, Dahm W J A and Dowlin D R 1997 *J. Fluid Mech.* **338** 127
[26] Parisi G and Frisch U 1985 *On the Singularity Structure of Fully Developed Turbulence, Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics*, North Holland, pp. 71–84
[27] Over T M and Gupta V K 1996 *J. Geophys. Res.* **101** 26319
[28] Davis A, Marshak A, Wiscombe W and Cahalan R 1994 *J. Geophys. Res.* **99** 8055
[29] Davis A, Marshak A, Wiscombe W and Cahalan R 1996 *Current Trends in Nonstationary Analysis* (1st edn.) (Singapore: World Scientific) pp. 97–158
[30] Mandelbrot B B 1997 *Fractals and Scaling in Finance* (1st edn.) (New York: Springer) pp. 174–176
[31] Tung W W, Moncrief M W and Gao J B 2004 *J. Climate* **17** 2736
[32] Gao J B and Rubin I 2001 *Comput. Commun.* **24** 1400
[33] Gao J B and Rubin I 2001 *Int. J. Commun. Syst.* **14** 783
[34] Peng C K, Buldyrev S V, Havlin S, Simons M, Stanley H E and Goldberger A L 1994 *Phys. Rev. E* **49** 1685
[35] Riley M A, Kuznetsov N, Bonnette S, Wallot S and Gao J B 2012 *Front. Physiol.* **3** 10
[36] Kuznetsov N, Bonnette S, Gao J B and Riley M A 2012 *Ann. Biomed. Eng.* **41** 1646
[37] Gao J B, Hu J, Mao X and Perc M 2012 *J. R. Soc. Interface* **9** 1956
[38] Gao J B, Hu J and Tung W W 2011 *PLoS ONE* **6** e24331