African Journal of Agricultural Research Vol. 6(14), pp. 3363-3368, 18 July, 2011
Available online at http://www.academicjournals.org/AJAR
DOI: 10.5897/AJAR11.497
ISSN 1991-637X ©2011 Academic Journals

Full Length Research Paper

Identification of weeds based on fractal dimension analysis of time series of weed leaf chlorophyll

Shuxi Cheng, Yongming Chen, Ping Lin and Yong He*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310058, China.

Accepted 15 June, 2011

A new application of fractal dimension analysis was initiated for identification of different kinds of broadleaf weeds. The distribution of chlorophyll time series of weed leaves exhibited self-similar geometrical characteristics. The fractal dimension analysis was conducted using Grassberger-Procaccia (G-P) phase space reconstruction algorithm. A total of 300 leaves of three weeds of *Oxalis corniculata* L. (OC), *Ixeris chinensis* (Thunb.) Nakai. (IC) and *Herba glechomae* L. (HG) (species) were sampled and analyzed. The correlation dimensions of time series of *O. corniculata*, *I. chinensis* and *H. glechomae* estimated by G-P algorithm are 8.050, 10.094 and 11.730, respectively. The distribution of chlorophyll time series was restricted in chaos environment and governed by strange attractors. The self-similar distribution property includes the information about weed varieties, which can be used for classification.

Key words: Fractal dimension, identification, chlorophylls, weed, Grassberger-procaccia (G-P) algorithm.

INTRODUCTION

Automatic identification of weeds is a very basic and important project in the areas of biosystems engineering. For the successful development of the weed identification systems, the precision mechanical and chemical weed control can be realized, which demonstrates great environmental and economic advantages. Varieties of related studies have been carried out on weed recognition (Alberto et al., 2008; Ishak et al., 2009; Wu et al., 2009). However it is rare to see the study on identification of weeds based on fractal dimension of time series of weed leaf chlorophyll. The theories of Fraction and Chaos theory (Sean, 2002; Takens, 1981) are used to investigate nonlinear problems such as dissipative dynamical system. These systems which show chaotic behavior have a strange attractor in phase space. Strange attractors are typically characterized by fractal dimension. There are multi-definitions of fractal dimension, where correlation dimension is sensitive to asymmetry of strange attractors of systems and able to reflect to dynamic structure strange attractors. Phase space reconstruction algorithm presented by Grassberger and Procaccia (1983) (G-P), is usually applied to measure the signal fractal dimension. The theory of Fraction is firstly introduced to investigate the chlorophyll time series of weed leaves in this paper. Chlorophyll values of three weeds, *Oxalis corniculata* L. *Ixeris chinensis* (Thunb.) Nakai and *Herba Glechomae* L. were measured by chlorophyll meter SPAD-502. SPAD unit can denote chlorophyll value. View the chlorophyll time series as complex dynamic systems, and use G-P algorithm to extract the fractal dimensions of dynamic systems. The objectives of this paper were to (1) study if there is self-similar distribution property of weed leaf chlorophyll and (2) identify the different kinds of weeds using fractal dimensions.

*Corresponding author. E-mail: yhe@zju.edu.cn. Tel: +86-571-86971143. Fax: +86-571-86971143.

Abbreviations: KNN, K nearest neighbour; **GP,** grassberger-procaccia; **SNV,** standard normal variate; **SV,** scale variate.

MATERIALS AND METHODS

Three broadleaf weeds, *O. corniculata* L., *I. chinensis* and *H. glechomae* used for study were located at Huajiachi Campus (Latitude: 30.273°, Longitude: 120.190°), Zhejiang University, Hangzhou, China. The samples used in the experiments consisted of 300 broadleaf weed leaves and each variety contained 100 leaves. The same group of sampled weed leaves all seemed to be the same age. The selected broadleaf weeds are widely distributed in Zhejiang Province. Their names and families are listed in Table 1.

Table 1. Three different kinds of broadleaf weed species[a].

Specimen	Name	Family
1	*Oxalis corniculata.*	Oxalidaceae
2	*Ixeris chinensis* (Thunb.) Nakai.	Compositae
3	*Herba glechomae* L.	Primulaceae

[a] The broadleaf weed species are referred to the encyclopaedia on the Baidu website at (http://baike.baidu.com/).

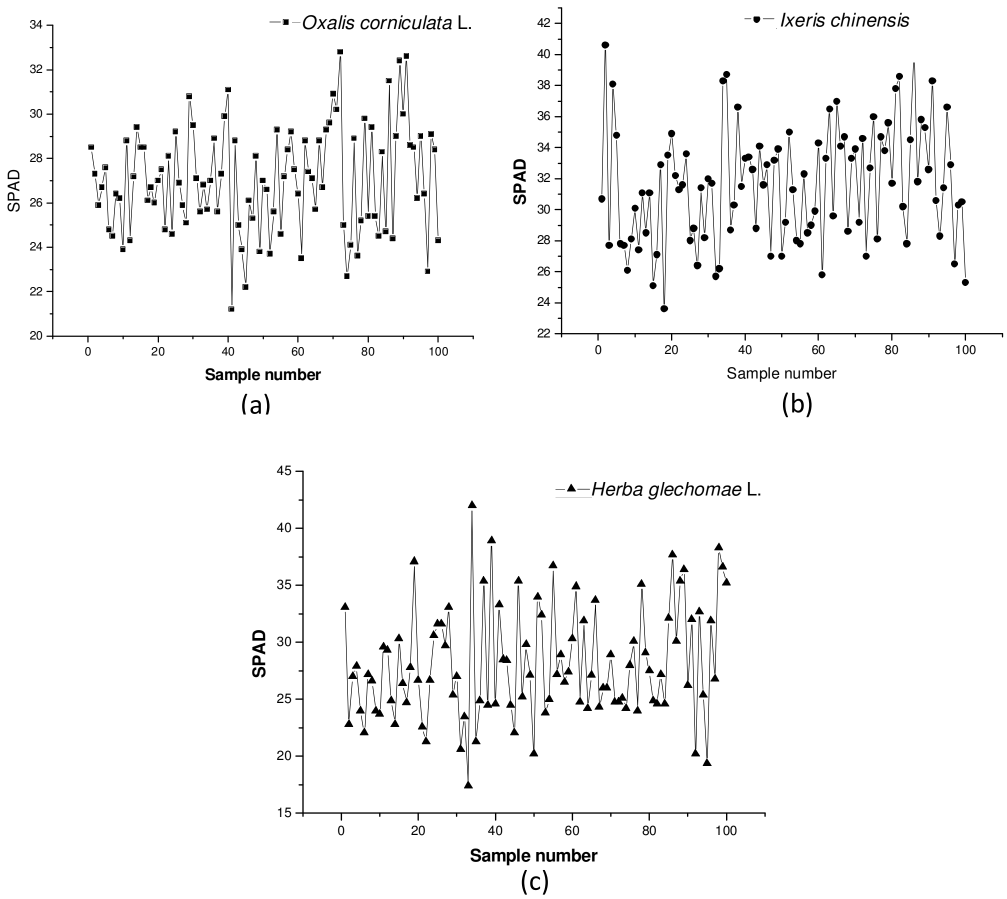

Figure 1. Chlorophyll or SPAD time series of weed leaves, *Oxalis corniculata* L. (OC), *Ixeris chinensis* (Thunb.) Nakai. (IC) and *Herba glechomae* L. (HG).

Devices and software

Lightweight handheld chlorophyll meter SPAD-502 (KONICA MINOLTA Inc.) was used to measure chlorophyll contents of weed leaves. Light source is 2LEDs (light-emitting diodes) and receptor is 1 silicon photodiode. Measurement system adopts the method of optical density difference at two wavelengths (around 650 and 940 nm, respectively). Measurements were taken by simply inserting a leaf and closing the measuring head. In this study, three replicate SPAD measurements were made at the centre of each weed leaf and get the average. The SPAD data were collected and on January 10, 2011. The status of weed leaf chlorophyll are widely affected by the climate, soil nutritional status, leaf phenology and age et al, However, the paper only dealt with the temporal status of foliar chlorophyll and all of the chlorophyll measurements were finished in one day. The sampled SPAD data of *O. corniculata*, *I. chinensis* and *H. glechomae* leaves are shown in Figure 1. The algorithms were developed and performed in Matlab R2009b (The MathWorks, Natick, and USA) in Windows.

Data preprocess

The chlorophyll time series of weed leaves can be viewed as a single time series composed of n elements, $X=\left[x_1,x_2,\cdots,x_n\right]^{\mathrm{T}}$. Data preprocess includes standard normal variate (SNV) and scale variate (SV) (Barnes et al., 1989).

SNV function:

$$y_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where μ is the mean, σ is the standard deviation. After SNV processing, the new time series is:

$$Y = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}.$$

Scale variate:

$$z_i = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)} \quad (2)$$

Where, $\min(y_i)$ is the minimum of $Y = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$. $\max(y_i)$ is the maximum of $Y = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$. After SV processing, the new time series is $Z = [z_1, z_2, \cdots, z_n]^{\mathrm{T}}$ which will be applied to input of phase space reconstruction.

Phase space reconstruction

The theory of phase space reconstruction was presented by Pacard and Takens (1981), who introduced chaos theory to analyze the nonlinear time series. If we want to identify whether there are chaos phenomenon in chlorophyll time series, we can reconstruct the phase space of dynamic system based on the time series. Chlorophyll time series is relatively complex, and can be viewed as a dynamic system. Its systematic evolutive dynamic function can be described by dynamic system control function:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = f_i(x_1, x_2, \cdots, x_m), \quad i = 1, 2, \cdots, m \quad (3)$$

Systematic time variables can be described as $m-\text{dimension}$ phase space tracks constructed by multi-variables $(x_1, x_2, \cdots, x_m)$:

$$X(t) = [x_1(t), x_2(t), \cdots, x_m(t)]^{\mathrm{T}} \quad (4)$$

For the single variable time series of chlorophyll, only need to transform Equation (3) to $m-\text{order}$ nonlinear differential equation:

$$x^{(m)} = F(x, x', x'', \cdots, x^{(m-1)}) \quad (5)$$

The new track of transformed chlorophyll time series is:

$$X(t) = [x, x', x'', \cdots, x^{(m-1)}]^{\mathrm{T}} \quad (6)$$

Discretize Equation (6), namely phase space reconstruction of chlorophyll time series of single variable $x_i (i = 1, 2, \cdots, m)$ delay time interval:

$$X(t) = [x(t), x(t+\tau), x(t+2\tau), \cdots, x(t+(m-1)\tau)]^{\mathrm{T}} \quad (7)$$

Where, τ is a fixed delay time interval, $\tau = k\Delta t, k = 1, 2, \cdots, N$, N is the number of phase points, $N = n-(m-1)\tau$, n is the length of the time series. $X(t)$ is m dimension phase space vector. Equation (7) describes the evolutive track of chlorophyll restriction series in $m-\text{dimension}$ phase space. The key technique of phase space reconstruction of dynamic system lies in selecting fixed delay time interval τ and embedding dimension m. In practice, τ and m are neither suitable large nor small and they restrict each other. Auto correlation function is used to estimate fixed delay time interval τ. G-P algorithm is used to select embedding dimension m.

Auto correlation function

If we need to reconstruct phase space, delay time interval τ must be solved. Auto correlation function (Fichera and Pagano, 2009) is usually used to select delay time interval τ. Suppose chlorophyll vector was $\{x_1, x_2, \cdots, x_n\}$ and its auto correlation function is denoted:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \mu)(x_{t+k} - \mu)}{\sum_{t=1}^{n}(x_t - \mu)^2} \quad (8)$$

Where r_k denotes $k-\text{order}$ order auto correlation coefficient. μ is mean. k is applied to delay time interval τ when correlation coefficient r_k approximates to 0.

Grassberger-procaccia (G-P) algorithm

The famous algorithm of correlation integral used to compute fractal dimension D whenever $D > 2$ was introduced by Grassberger and Procaccia (1983). The definition of the correlation integral is:

$$C_m(r) = \frac{1}{N^2}\sum_{i,j=1}^{N} H(r - \|X_i - X_j\|) \quad (9)$$

Where, $\|X_i - X_j\|$ denotes Euclidean length between X_i and X_j. $H(\bullet)$ is the Heaviside function:

$$H(\bullet) = \begin{cases} 1, & r > \|X_i - X_j\| \\ 0, & r \le \|X_i - X_j\| \end{cases} \quad (10)$$

$C_m(r)$ is the probability for the distances of $m-\text{dimension}$ phase space points smaller than r. There are proportional relationships between $C_m(r)$ and r in certain section:

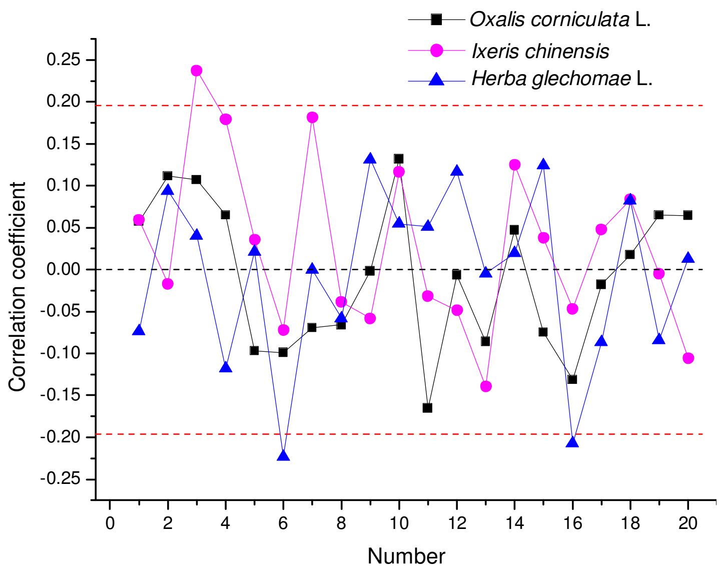

Figure 2. Auto correlation coefficients of chlorophyll time series of OC, IC and HG leaves.

$$C_m(r) \propto r^{D(m)} \tag{11}$$

Where, $D(m)$ is correlation dimension. Equation (11) is called scale relationship and can be transformed to:

$$D(m) = \lim_{r \to 0} \frac{\ln C_m(r)}{\ln r} \tag{12}$$

When m is enough large and do not change with $D(m)$, the staturation correlation dimension of chlorophyll restriction series can be obtained, namely fractal dimension of strange attractor D_2:

$$D_2 = \lim_{m \to \infty} D(m) \tag{13}$$

Where, D_2 corresponds to the minimal embedding dimension m^*, and the dimension of phase space is solved. The dynamical information of the chlorophyll time series of weed leaves can be described by D_2. The existence of saturation correlation dimension accounts for the existence of systematic strange attractor, which indicates that there is chaos characterization in system. On the contrary, if system is random or determinate system, the size of saturation correlation dimension denotes the complex of systematic evolution.

RESULTS AND DISCUSSION

Compute delay time interval

Equation (8) was applied to auto correlation estimation of SPAD time series of *O. corniculata*, *I. chinensis* and *H. glechomae* leaves. The estimation results were showed in Figure 2. We selected corresponding delay time intervals of *O. corniculata*, *I. chinensis* and *H. glechomae* $\tau =$ 1, 2 and 3, respectively.

Compute saturation correlation dimension D_2

Equations (9 to13) were used to compute saturation correlation dimension D_2 of chlorophyll time series of *O. corniculata*, *I. chinensis* and *H. glechomae* leaves. The computing results were shown in Figure 3. Minimal embedding dimension m^* and saturation correlation dimension D_2 of the chlorophyll time series of these three classes of weed leaves were presented in Table 2. In Figure 3a when delay time interval τ of *O. corniculata* was obtained, the curve relationships of $\log_e C_m(r)$ and $\log_e r$ in scale-free region approximated to linear. With the increment of embedding dimension m, linear curves trended to parallel. The slopes, namely $D(m)$, of $\log_e C_m(r) \sim \log_e r$ in scale-free regions were computed using least square method. From the change of the estimating correlation dimension $D(m) \sim m$, when $m \geq 12$, $D(m)$ tend to stable.

Thus we convinced that saturation correlation dimension $D_2 = 8.050$ and minimal embedding dimension $m^* = 12$, that is, movement tracks of phase space in which the change of dynamic systems of chlorophyll time series of *O. corniculata* leaves contracted to about 8.050-dimension strange attractor. Twelve

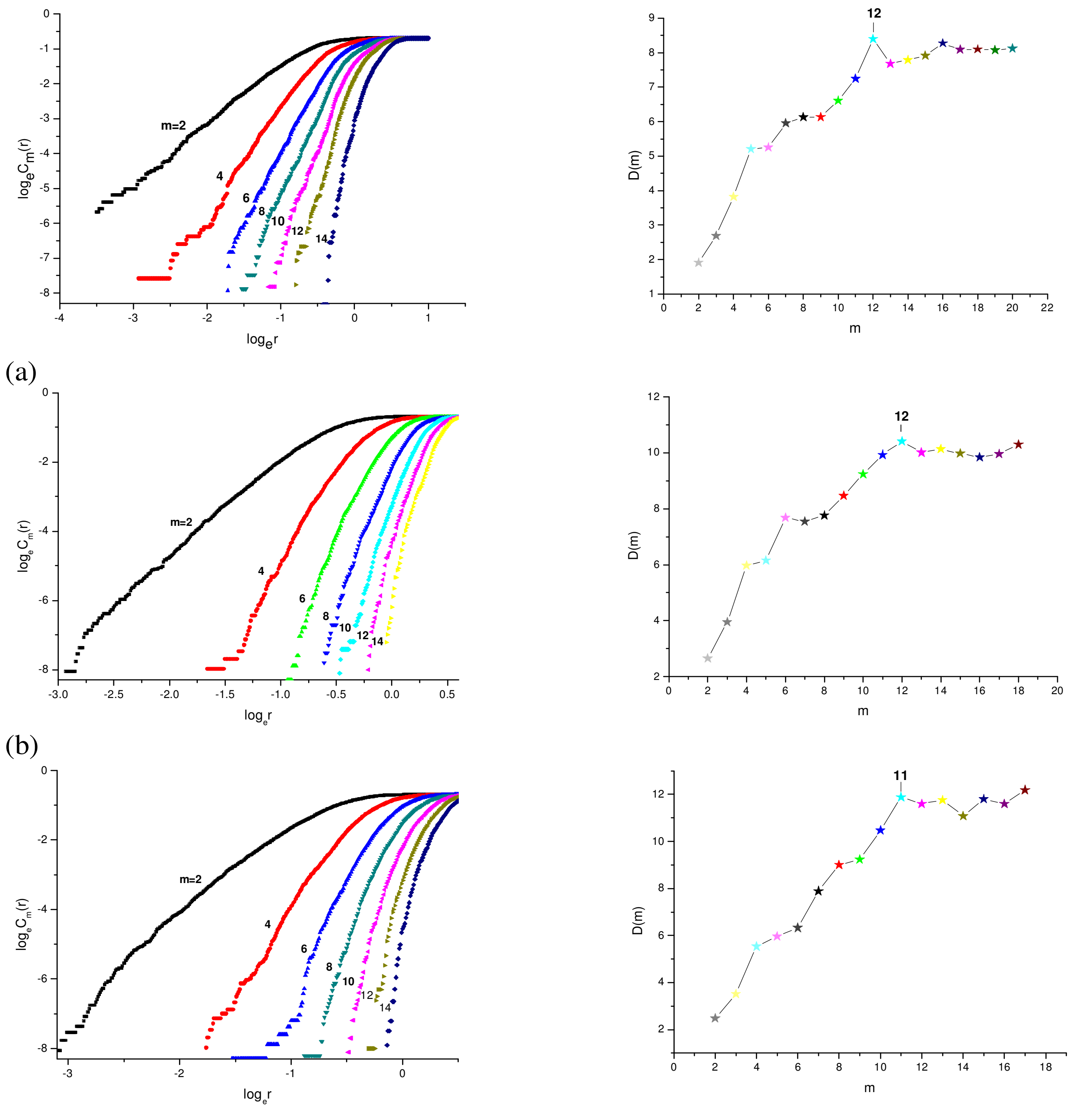

Figure 3. The relationships of $\log_e C_m(r) \sim \log_e r$ (left) and $D(m) \sim m$ (right) of (a) OC, (b) IC and (c) HG.

independent variables could describe the dynamic system. There was chaos phenomenon in chlorophyll time series of *O. corniculata* leaves. The analytical methods of time series of *I. chinensis* and *H. glechomae* leaf chlorophyll were the same as the *O. corniculata*. As shown in Figure 3b and c, the saturation correlation dimensions D_2 of *I. chinensis* and *H. glechomae* were 10.094 and 11.730, respectively, and the least embedding dimension m^* were 12 and 11.

There were chaos phenomena in time series of *I. chinensis* and *H. glechomae* leaf chlorophyll.

Analyzing and identifying weed leaf chlorophylls

The SPAD or chlorophyll time series of weed leaves of *O.*

corniculata, *I. chinensis* and *H. glechomae* could be viewed as a nonlinear dynamic system. The theories of Fraction and Chaos were carried out to analyze their evolutive process. The above experimental results showed that the chlorophyll time series of *O. corniculata*,

Table 2. The calculation results of time increment τ , minimal embedding dimension m^* and saturation dimension D_2 .

Name	Delay time τ	Embedding dimension m^*	Correlation dimension D_2
Oxalis corniculata.	1	12	8.050
Ixeris chinensis (Thunb.) Nakai.	2	12	10.094
Herba glechomae L.	3	11	11.730

I. chinensis and *H. glechomae* leaves had different fractal dimensions, 8.050, 10.094 and 11.730, respectively. It accounted that there were self-similarities of time series of weed leaf chlorophyll, namely there was chaos status in time series of weed leaf chlorophyll. It also showed that chlorophyll time series of different weeds were governed by different strange attractors. From the above theory of phase space reconstruction, it was known that if there were chaos phenomena in a certain nonlinear dynamic system and its dimensions for analysis are enough large, the G-P algorithm could be used to determine the fractal dimension of the system.

The computed fractal dimension represented the intrinsic characteristic of the nonlinear system and was not influenced by the changes of the computed dimensions of the nonlinear system, In order to understand if a certain time series of weed leaf chlorophyll was governed by the strange attractors, different weed leaves did not need to be measured numerous times, and the measuring times could be approximately estimated using the known delay time interval τ and minimal embedding dimension m^* (Table 1). The status of weed leaf chlorophyll might be widely affected by the climate, soil nutritional status, leaf phenology and age et al. and there were system and stochastic errors in the measured SPAD series, the saturation correlation dimensions of the measured samples D_2' were not equal to the models' D_2. In order to identify the unknown samples, the differences between of the saturation correlation dimensions between the samples and the models were computed. The samples and the model with the smallest difference value were considered as the same group. The method was equal to the case of the k nearest neighbour (KNN) classifier using $k=1$. Of course, those weeds used here were not representative of all weeds in fields and more different kinds of the weed needed further investigation.

Conclusions

A novel application of fractal dimension analysis was initiated for the identification of three different kinds of weeds. The paper only dealt with the temporal status of foliar chlorophyll and did not considered the other influence on weed foliar chlorophyll for example, the climate, soil nutritional status, and leaf phenology and age, however these factors were significant for further investigation. Through the analysis of chlorophyll time series of *O. corniculata*, *I. chinensis* and *H. glechomae* leaves, it showed that there were chaos phenomena in chlorophyll time series, which implied that the complex and diverse chlorophyll time series of weed leaves might be the evolutive results of nonlinear chaos dynamic systems. The distributions of weed leaf chlorophyll, *O. corniculata*, *I. chinensis* and *H. glechomae* demonstrated different similarity. The self-similarities could be estimated by G-P algorithm and used for the identification of different kinds of weed species. However, those weeds used here were not representative of all weeds in fields and more different kinds of weeds needed further investigation.

ACKNOWLEDGMENTS

This study was supported by Natural Science Foundation of China (31071332), Zhejiang Provincial Natural Science Foundation of China (Z3090295), Important Zhejiang Provincial Science and Technology Specific Projects (2009C12002), National Special Public Sector Research of Agriculture (200903044) and National Agricultural Science and Technology Achievements Transformation Fund Programs (2009GB23600517).

REFERENCES

Alberto T, Xavier PB, Gonzalo P, Angela R, César FQ (2008). A new vision-based approach to differential spraying in precision agriculture. Comput. Electron. Agr., 60: 144-155.
Barnes R, Dhanoa M, Lister J (1989). Standard normal variable transformation and detrending of near infrared diffuse reflectance spectra. Appl. Spectrosc., 43: 772-777.
Ficera A, Pagano A (2009). Monitoring combustion unstable dynamics by means of control charts. Appl. Energ., 86: 1574-1581.
Grassberger P, Procaccia I (1983). Characterization of strange attractors. Phys. Rev. Lett., 50(5): 346-349.
Ishak AJ, Hussain A, Mustafa MM (2009). Weed image classification using Gabor wavelet and gradient field distribution. Comput. Electron. Agr., 66: 53-61.
Sean PR (2002). Fractal theory for the compensation effect observed in

a surface diffusion process studied using deuteron NMR. Langmuir, 18(5): 1613-1618.
Takens F (1981). Detecting strange attractors in turbulence. Lect. Notes Math., (eds. D.A. Rand and L. - S. Young). New York: Springer, pp. 366-381.
Wu LL, Wen YX (2009). Weed/corn seedling recognition by support vector machine using texture feat ures. Afr. J. Agric. Res., 4(9): 840-846.