



Quantitative effects of composting state variables on C/N ratio through GA-aided multivariate analysis

Wei Sun^a, Guo H. Huang^{a,b,*}, Guangming Zeng^c, Xiaosheng Qin^d, Hui Yu^a

^a Faculty of Engineering and Applied Science, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2

^b MOE Key Laboratory of Regional Energy Systems Optimization, Sino-Canada Energy and Environmental Research Academy, North China Electric Power University, Beijing, 102206 China

^c MOE Key Laboratory of Environmental Biology and Pollution Control, College of Environmental Science and Engineering, Hunan University, Changsha, Hunan, 410082, China

^d School of Civil and Environmental Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798, Singapore

ARTICLE INFO

Article history:

Received 20 August 2010

Received in revised form 1 December 2010

Accepted 17 December 2010

Available online 22 January 2011

Keywords:

Food waste composting

Stepwise cluster analysis

Genetic algorithm

Variable selection

The C/N ratio

ABSTRACT

It is widely known that variation of the C/N ratio is dependent on many state variables during composting processes. This study attempted to develop a genetic algorithm aided stepwise cluster analysis (GASCA) method to describe the nonlinear relationships between the selected state variables and the C/N ratio in food waste composting. The experimental data from six bench-scale composting reactors were used to demonstrate the applicability of GASCA. Within the GASCA framework, GA searched optimal sets of both specified state variables and SCA's internal parameters; SCA established statistical nonlinear relationships between state variables and the C/N ratio; to avoid unnecessary and time-consuming calculation, a proxy table was introduced to save around 70% computational efforts. The obtained GASCA cluster trees had smaller sizes and higher prediction accuracy than the conventional SCA trees. Based on the optimal GASCA tree, the effects of the GA-selected state variables on the C/N ratio were ranged in a descending order as: $\text{NH}_4^+\text{-N}$ concentration > Moisture content > Ash Content > Mean Temperature > Mesophilic bacteria biomass. Such a rank implied that the variation of ammonium nitrogen concentration, the associated temperature and the moisture conditions, the total loss of both organic matters and available mineral constituents, and the mesophilic bacteria activity, were critical factors affecting the C/N ratio during the investigated food waste composting. This first application of GASCA to composting modelling indicated that more direct search algorithms could be coupled with SCA or other multivariate analysis methods to analyze complicated relationships during composting and many other environmental processes.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Efficient operations of composting relies on insights of relationships between state variables (e.g. oxygen concentration, ash content, moisture content, and pH) and specific characteristics (e.g. microbial activities, maturity, and stability). Previously, many experimental approaches were developed in analyzing these relationships (Chefetz et al., 1996; Tiquia et al., 1996; Herrmann and Shann, 1997; Hsu and Lo, 1999; Sanchez-Monedero et al., 2001; Liang et al., 2003; Galvez-Sola et al., 2010). However, experiment-based evaluations could hardly help quantify the interactions among multiple composting state variables. In comparison, a model-based analysis could help examine the inherent impacts of various factors on the biological and physiochemical processes and gain an in-depth insight into the related mechanisms (Hamoda et al., 1998; Turner et al., 2005; Mason, 2006; Chikae et al., 2006, 2007; Lin et al., 2008; D'Imporzano et al., 2008; Khalil et al., 2008; Mudhoo and Mohee, 2008; Vlyssides et al., 2009; Giusti and Marsili-

Libelli, 2010). Nevertheless, dynamic, nonlinear and interactive features of these physicochemical or biological properties in composting inherently make the analysis as a challenge. Conventional continuous and linear methods cannot efficiently reflect such complicated relationships.

Stepwise cluster analysis (SCA) is an emerging non-parameter regression technology. It includes a series of cut or merge operations according to given statistic criteria and finally generates a cluster tree in the sense of probability. SCA has been applied to predicting air quality in the urban environment (Huang, 1992), supporting diagnosis of lung cancer (Ren et al., 1997), and optimizing groundwater bioremediation processes (Qin et al., 2007; He et al., 2008b). However, applications of SCA in approximating the relationships between state variables and specific characteristics during composting processes were relatively limited (Sun et al., 2009). Meanwhile, one weakness of SCA is that its performance is sensitive to both input variables and internal parameters. Effective identification of these variables and parameters will be beneficial for improving the accuracy of SCA predictability. The identification process essentially involves solution of a nonlinear, discrete optimization model with the objective function being the performance of SCA; this could hardly be handled by conventional

* Corresponding author. Faculty of Engineering, University of Regina, Regina, Saskatchewan, Canada, S4S 0A2. Tel.: +1 306 585 5631; fax: +1 306 585 4855.

E-mail address: huangg@iseis.org (G.H. Huang).

gradient-based methods. Genetic algorithm (GA) is a powerful searching technique in seeking optimal solutions in large and complex decision spaces through imitating the principle of natural evolution. It has been successfully applied to variable or factor selection in various statistical predictions or machine learning studies (Depczynski et al., 2000; Cavill et al., 2009), as well as detection of the most relevant spectral region related to the composting process (Martinez-Sabater et al., 2009).

Therefore, the objective of this study is to develop a GA-aided stepwise-cluster analysis model (GASCA) through integrating GA and SCA within a general framework to quantify the nonlinear relationships between state variables (time, temperatures, pH, moisture content, ash content, organic content, NH_4^+ -N concentration, cumulative NH_3 emissions and bacterial biomass) and the C/N ratio in food waste composting. It entails: screening of significant state variables and system parameters in virtue of the powerful search ability of GA; development of a GA-aided stepwise-cluster model for analyzing the effects of multiple state variables on the C/N ratio; and demonstration of the proposed method based on the data from bench-scale composting experimental systems.

2. Materials and methods

2.1. Feedstock and reactors

We chose synthetic food waste as the composting feedstock and used six runs of designed composting experiments (bench-scale reactors) in our laboratory to describe various representative operation conditions. The synthetic food waste was produced using potato, carrot, ground pork, steamed rice (food products from a local grocery store) and American Elm leaves. Cooked soybean was added in Runs 1 and 2 to lower the C/N ratio. Baking soda (Sodium hydrogen carbonate) was added in Runs 3 and 4 to adjust the pH value. Coal ash was added in Run 6 as a bulking agent. These materials were cut into pieces of approximately 5 mm in size and mixed thoroughly. The composting mixtures were manually turned over with a shovel once a day, and three samples (less than 50 g in total) from three locations of each reactor were collected for analyzing pH, moisture content, ash content, organic content, NH_4^+ -N concentrations, bacterial biomass and C/N ratio. The weights of samples were kept minimal to avoid their effects on the composting process. The compositions and physiochemical characteristics of initial composting mixtures for the six runs are listed in Table 1.

Each reactor is a cylindrical PVC tube (40.64 cm high and an effective volume of approximately 30 l) with three layers of heat-insulating materials preventing heat loss. On the bottom, an air inlet and a leachate outlet were connected. The inlet air, blown in by a vacuum pump, was distributed flowing upward through a holed vinyl pipe in the aeration distribution chamber near the bottom of the reactor. The aeration rate was monitored by a flow-meter. On the top, an air outlet, a thermometer point and a sampling port were distributed. The outlet air was discharged through a vinyl tube into a wide mouth flask containing H_2SO_4 solution (100 ml 0.1 M) to absorb NH_3 . The final gas passing through a condensation bottle was discharged to the lab ventilation system. A flow diagram of the composting reactor is shown in Fig. 1.

2.2. State variables and C/N ratio

With the ambient temperature being maintained at 20 ± 1 °C for all experimental runs, we placed a stainless steel rod with two thermometers (Traceable 15-077-9E, Control Company, TX, USA) at different depths of the compost matrix. The lower thermometer was fixed at a location 2.54 cm away from the segmentation plate to obtain the surface's temperature. The upper one was placed in the centre of the compost matrix and adjusted everyday with the change of compost height to acquire the core's temperature. The pH in a solution of compost and water (weight ratio: 1:2) was measured through a bench top pH/Temperature Meter (Thermo Orion, 410A Plus) (Thomas, 1996). The moisture content by weight loss of compost sample was determined by the gravimetric method (Gardner, 1986); the ash content and the organic content were measured by the ignition method (Nelson and Sommers, 1982). In addition, NH_4^+ -N concentrations (mg/kg, wet Sample) and cumulative NH_3 emissions ($\mu\text{g}/\text{day}$) were analyzed through the FIAStar 5000 Analyzer (Foss analytic AB, Sweden) equipped with 5027 Sampler, interference filters (540 and 720 nm), and reduction columns (Foss Analytical AB, 2001). The colony numbers of thermophilic and mesophilic bacteria were counted by the spread plate counting method (Wollum, 1982; Germida, 1993). The oxygen concentrations in the surrounding air and the exhaust gas of the reactor were monitored every 3 h by a M40 Multi-Gas Monitor (Industrial Scientific Corp., Oakdale, PA, USA), and the C/N ratio (the ratio of total carbon and total nitrogen) was measured by the LECO TruSpec CN Determinator (LECO Corporation, St. Joseph, MI, USA). More experimental details can be found in the work of An (2006).

Table 1
Initial state of raw composting mixtures for six barrels.

	Barrel	Unit	1	2	3	4	5	6
Substrate	Potato	kg	2	2	2.2	2.2	2.0	1.7
	Steamed rice	kg	3.1	3.1	3.5	3.5	3.2	2.8
	Carrot	kg	3	3	3.4	3.4	3.1	2.7
	Leaves	kg	0.7	0.7	0.8	0.8	0.7	0.6
	Meat	kg	0.5	0.5	0.6	0.6	0.5	0.5
	Soy bean	kg	3	6	–	–	–	–
	Starting culture	kg	0.5	0.5	0.5	0.5	0.5	0.5
	Coal ash	kg	–	–	–	–	–	4.0
	Baking soda	kg	–	–	0.2	0.3	–	–
	Total weight	kg	12.8	15.8	11.2	11.3	12.8	15.8
Chemical composition	pH	–	5.86	5.96	8.85	8.77	5.89	5.97
	Moisture content	%	56.98	56.38	62.55	66.65	64.27	50.09
	Ash content	%	1.88	2.08	2.90	2.38	1.48	29.73
	Organic content	%	41.14	41.54	34.55	30.97	34.25	20.18
	Carbon (dry)	%	47.93	47.77	43.67	44.37	45.88	19.70
	Nitrogen (dry)	%	3.25	3.17	1.34	1.48	1.44	0.56
	C/N ratio	–	14.75	15.08	32.63	29.89	31.9	34.99
Microorganism	Thermophilic bacteria	log CFU/g (dry)	4.667	7.146	4.728	4.477	6.080	5.478
	Mesophilic bacteria	log CFU/g (dry)	12.366	11.341	12.284	11.322	8.447	13.068

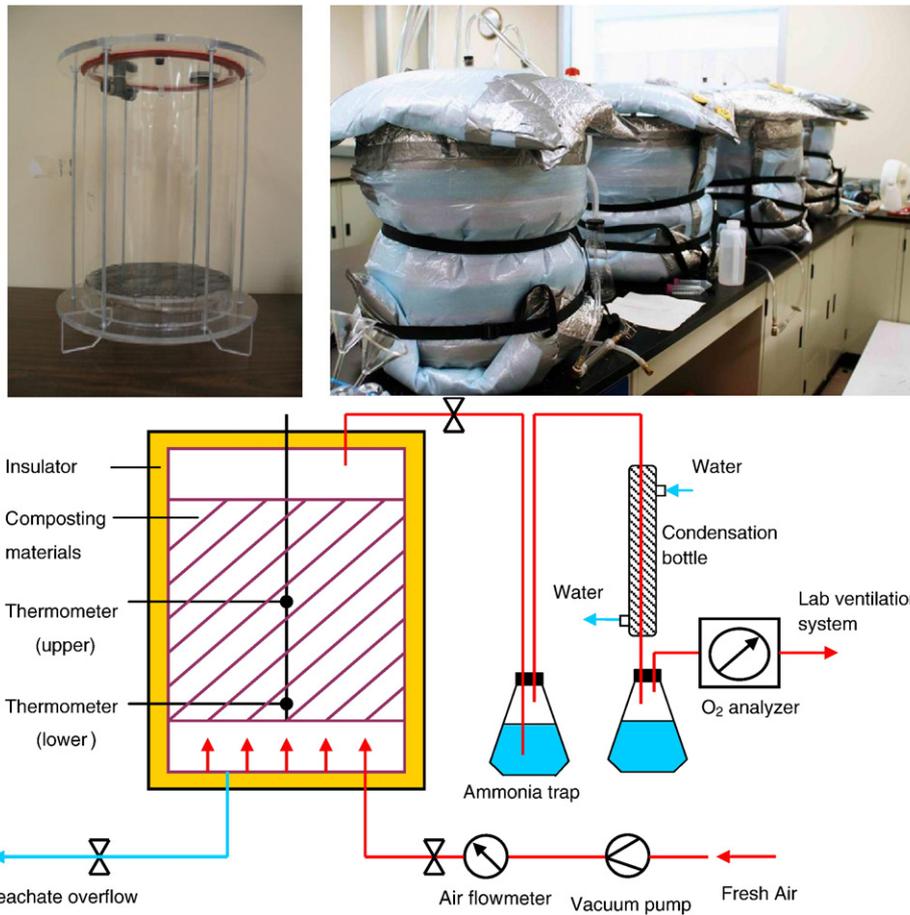


Fig. 1. Schematic diagram of the composting system.

3. Theoretical

3.1. Framework of GASCA

A GA-aided stepwise cluster analysis for describing the relationship between state variables and the C/N ratio during food waste composting consists of several major steps (Fig. 2). Firstly, all the observed variables during the six runs of composting experiments are collected as the dataset for the GASCA model, since the sample numbers from a single experimental run are not enough for constructing GASCA models. Then, GA searches optimal sets of specified state variables and SCA parameters through a number of

standard procedures such as initiation, crossover, and mutation. An index based on both the normalized root mean squared error (RMSE) of the training set and the SCA tree's size is used as fitness function for evaluating the performance of each potential set of parameters (i.e. chromosomes). Meanwhile, a proxy table will be introduced to store the SCA characteristics (the chromosome information and its corresponding fitness) so as to avoid unnecessary and repetitive training of SCA. GA will continue searching until the termination criteria are satisfied. Finally, the optimal combination of parameters is identified and further used in SCA for approximating the relationship between the selected state variables and the C/N ratio.

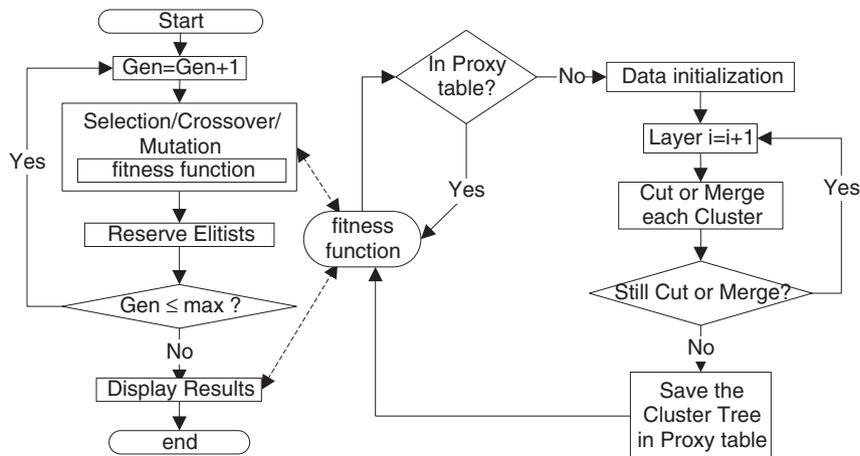


Fig. 2. Flow chart of GASCA.

3.2. Dataset

All the observed data from the six runs of food waste composting were collected (198 samples) and combined to form the whole dataset. Each sample was a combination of inputs (i.e. state variables and SCA's inherent parameters) and the C/N ratio. The state variables had potential effects on C/N ratio variation; they included time (X_1), mean temperature (X_2), pH (X_3), moisture content (X_4), ash content (X_5), organic content (X_6), NH_4^+ -N concentration (X_7), cumulative NH_3 emissions (X_8), log colony count of thermophilic bacteria (X_9), log colony count of mesophilic bacteria (X_{10}), upper temperature (X_{11}), and lower temperature (X_{12}). Although the sum of X_4 , X_5 and X_6 was one, the three state variables were still included since it was difficult to exclude one of them from the candidates subjectively; The SCA's inherent parameters were α_{cut} (significance level for cutting clusters), α_{merge} (significance level for merging clusters), and N_{min} (the minimum number of samples within tip clusters). The C/N ratio was the concerned output. The entire dataset was randomly divided into a training set (167 samples, 84%) and a test set (31 samples, 16%). The training set was used for calibrating the GASCA (or SCA) forecasting trees and the test one was for verification in independent samples from different reactors. Since the magnitudes of state variables were significantly different from each other, we firstly converted all input datasets into the range of $[-1, 1]$ before the training process and then transformed the values of the predicted dependent variables back to their original ranges after the training was completed.

3.3. Principle of stepwise cluster analysis (SCA)

SCA is based on the theory of multivariate analysis of variance. Generally, two phases are involved in its implementation: training and test. In the training phase, the sample dataset is cut into two subsets (clusters) and then two are merged into one during the iterative training process. According to Wilks' likelihood-ratio criterion, the cutting point is optimal only if the value of Wilks' Λ is minimum (Wilks, 1962). Since the Λ is directly related to the F statistic, the sample means of the two sub-clusters can be compared for significant differences through F test (Rao, 1952). Thus, the criteria of cutting (or merging or not) clusters rely on a set of F tests (Rao, 1965; Tatsuoka, 1971). Based on the F tests, a tree is shaped step by step until no subsets could be cut or merged any more. In the test phase, the values of the new independent variables in the test samples will be used as references to determine which child cluster a sample in the parent cluster will enter. Thus, the sample would gradually enter one tip cluster. The mean of dependent variables from the training set at the tip cluster that the sample finally enters becomes the predicted value of the sample's dependent variable. A more detailed description of SCA can be referred to our previous work (Qin et al., 2008).

3.4. Variable/parameter selection using GA

An integrated GASCA model combining GA with SCA would possess abilities in both variable searching and nonlinear fitting. GA transfers chromosome information (certain independent variables and inherent parameters) to SCA and then SCA returned the fitness (an index reflecting both structure and performance of the SCA tree) to the corresponding chromosome in GA. In detail, the chromosome is a 19-bit binary string. The first twelve binary strings encode all the SCA's candidate independent variables, where the '1' represents that its corresponding independent variable is selected while the '0' means not. The last seven binary strings represent inherent parameters of SCA (α_{cut} , α_{merge} and N_{min}). In order to unify the calculation, we chosen several representative values within the appropriate ranges of the three parameters respectively and used binary strings to represent the locations of each value in the ranges. Thus the binary strings are capable of describing real-number values of SCA's inherent parameters. The mapping relationships between the binary string and the variables/parameters are illustrated in Fig. 3.

An index based on both the SCA's fitting accuracy and the SCA tree's complexity was developed as the fitness function. The accuracy of SCA's fitting is defined as RMSE of the training set. The complexity of SCA tree is depicted as the sum of normalized layer number of the SCA ($N_{s,\text{layer}}$), the normalized node number of the SCA ($N_{s,\text{node}}$) and the normalized number of independent variables ($N_{s,x}$). Four weighting coefficients (μ_1 , μ_2 , μ_3 , and μ_4) were introduced to balance the accuracy and the complexity. Thus the fitness function has the following definition:

$$\text{Fitness} = \mu_1 \text{RMSE}_s + \mu_2 N_{s,\text{layer}} + \mu_3 N_{s,\text{node}} + \mu_4 N_{s,x} \tag{1}$$

$$\text{RMSE}_s = \frac{\sqrt{\frac{\sum_{j=1}^n (y_j - y_{\text{training},j})^2}{n-1}} - \text{RMSE}_{\text{min}}}{\text{RMSE}_{\text{max}} - \text{RMSE}_{\text{min}}} \tag{2}$$

$$N_{s,\text{layer}} = \frac{N_{\text{layer}} - N_{\text{layer, min}}}{N_{\text{layer, max}} - N_{\text{layer, min}}} \tag{3}$$

$$N_{s,\text{node}} = \frac{N_{\text{node}} - N_{\text{node, min}}}{N_{\text{node, max}} - N_{\text{node, min}}} \tag{4}$$

$$N_{s,x} = \frac{N_x - N_{x, \text{min}}}{N_{x, \text{max}} - N_{x, \text{min}}} \tag{5}$$

where RMSE_{min} and RMSE_{max} are ideal minimum and maximum values of RMSE, respectively; n is the number of samples in the training set; y_j and $y_{\text{training},j}$ are the predicted and observed values for C/N in the j th sample in the training set, respectively; N_{layer} is the layer number of the SCA tree. $N_{\text{layer, min}}$ and $N_{\text{layer, max}}$ are ideal minimum and maximum values of N_{layer} .

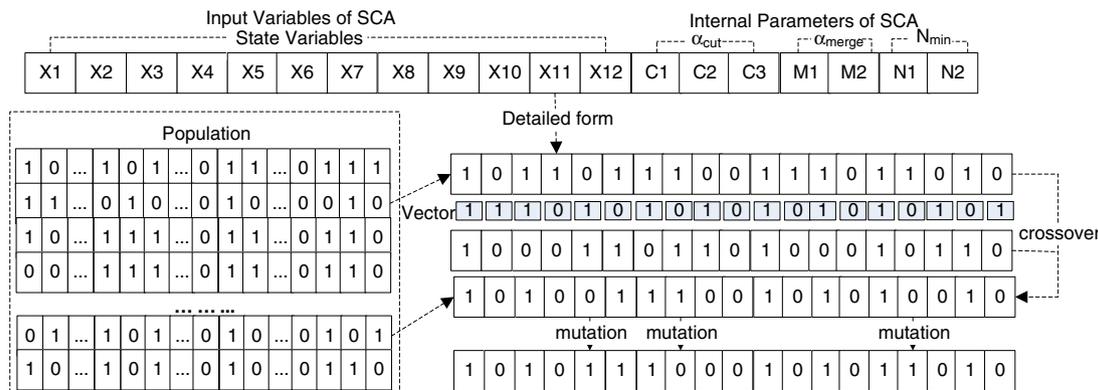


Fig. 3. Principle of GASCA.

respectively; N_{node} is the node number of the SCA tree; $N_{\text{node,min}}$ and $N_{\text{node,max}}$ are ideal minimum and maximum values of N_{node} , respectively; N_x is the number of independent variables; $N_{x,\text{min}}$ and $N_{x,\text{max}}$ are ideal minimum and maximum values of N_x , respectively.

Similar to the construction of a single chromosome, a population with a group of chromosomes in GA would be randomly generated. Parents would be selected based on individuals' fitness values using the stochastic uniform method, then further intercrossed to exchange their genetic information by the scattered crossover method, and mutated through changing their information randomly by the Gaussian mutation method (The MathWorks, 2008). Generation by generation, parents with worse fitness would be replaced by children with better fitness in the evaluated population. Once the stopping criterion was satisfied; the chromosome with the lowest fitness value would represent the information of the final optimal combination of variables/parameters for SCA.

3.5. Program implementation of GASCA

A flow diagram of GASCA is presented in Fig. 2. All programs were written in Matlab R2008a (The MathWorks, 2008). The main program for coupling GA and SCA is clarified as follows:

Step 1: Before SCA training, the entire dataset (the state variables and C/N) is scaled to fall within the range of -1 to 1 ;

Step 2: A number (the default is 100) of randomly generated chromosomes (of a bit length equal to the number of state variables and inherent parameters chosen) form a population;

Step 3: GA invokes the SCA function to evaluate fitness of every chromosome in the population;

Step 4: New chromosomes are generated through selection, crossover, and mutation operations; GA invokes the SCA function to calculate fitness values of new chromosomes.

Step 5: Replace the old chromosomes with new chromosomes;

Step 6: After all operations, the chromosome with the worst fitness of the entire population is replaced with the current elitist;

Step 7: Return to Step 4 to begin the evolution of the next generation; if the generation number exceeds a certain value (e.g. 25), the program will end.

The computational procedures of the SCA function are described as follows:

Step 1: GA will search the proxy table to find out whether the same chromosome is calculated already or not; if yes, the corresponding fitness will be assigned to the chromosome directly and the program goes to Step 12, and if no, goes to Step 2;

Step 2: GA invokes SCA to evaluate the fitness of every chromosome: the entire training dataset will be cut into a smaller one based on the independent variable information embedded in the chromosome; such a dataset is treated as a leaf cluster (i.e. a cluster without sub-clusters); the parameter information (α_{cutting} , α_{merging} and N_{min}) will be used in Steps 9, 11 and 12;

Step 3: For every leaf cluster, each state variable X_{id} (i.e. independent variables, $id = 1, 2, \dots, m$) is chosen in the order of X_1 to X_m ; the samples (including both independent and dependent variables) in cluster σ (matrix) are sequenced to $\sigma(X_{id})$ in an ascending order according to the value of X_{id} ;

Step 4: For the matrix $\sigma(X_{id})$, every cutting line nr (between row i to row $i + 1$, $nr = 1, 2, \dots, n - 1$) is chosen in the order from 1 to $n - 1$; n is the number of samples for cluster $\sigma(X_{id})$; the cluster $\sigma(X_{id})$ will be divided into two sub-clusters β and γ according to the cutting line;

Step 5: The Wilks' Λ based on the dependent variables (Y_β and Y_γ) is calculated; by comparing all $m^*(n - 1)$ values of Wilks' Λ , the minimal value of $\Lambda(id^*, nr^*)$ can be obtained;

Step 6: Calculate the F_{cal} value; if $F_{\text{cal}} \geq F_{\alpha\text{-cutting}}$, the parent cluster can be cut into two sub-clusters;

Step 7: The independent variable and the value in the cutting point (x_{id^*} , nr^*) becomes a judging criterion; all samples in this cluster with x_{id^*} , $nr \leq x_{id^*}$, nr^* are allocated to one sub-cluster, with the rest to the other;

Step 8: Repeat Steps 3 to 7 until all clusters are tested for possible cutting in the present layer;

Step 9: When the cutting operation in the present layer is done, the merging operations begin; if other old leaf clusters exist, these new leaf clusters will be tested along with all other leaf clusters to demonstrate whether pairs of them could be merged into one cluster; the Wilks' Λ and F_{cal} values will be calculated for each pair clusters; if $F_{\text{cal}} \geq F_{\alpha\text{-merging}}$, the two clusters will be considered having no significant difference and can be merged;

Step 10: If there are no other old leaf clusters or no merged operations are needed, the new leaf clusters will be treated as new clusters and tested for further cutting or merging operations according to Steps 3 to 9;

Step 11: If no leaf cluster can be cut or merged or the number of samples in every cluster is less than N_{min} , the training process is completed with the SCA tree being shaped; then, the fitness of chromosome is calculated based on Eqs. (1)–(5);

Step 12: Return to Step 1 for evaluating the next chromosome; if the fitness of each existing chromosome has been evaluated, return to the main program.

4. Results

4.1. Structure of GA

In general, GA's fitness function and operation (selection, crossover, and mutation) configurations have significant impacts on its performance. The fitness function is always problem-oriented; determination of its value (Eqs. (1)–(5)) depends on the relevant coefficients selection, according to the datasets of food waste composting. Through error and test, four weighting coefficients (i.e. μ_1 , μ_2 , μ_3 , and μ_4) for balancing the accuracy and the complexity of SCA were determined as 1.04, 0.35, 0.37 and 0.024. The coefficients for normalizing SCA' complexity calculation (i.e. RMSE_{min} , RMSE_{max} , $N_{\text{layer,min}}$, $N_{\text{layer,max}}$, $N_{\text{node,min}}$, $N_{\text{node,max}}$, $N_{x,\text{min}}$ and $N_{x,\text{max}}$) were configured as 0, 10, 3, 150, 3, 1500, 2 and 12, respectively. GA's standard operations are controlled by the relevant parameters, including the population size (PS), the selection rate (SR), the crossover fraction (CF, or crossover rate), and the mutation rate (MR). Usually, the larger the PS, the higher the CF; the higher the MR, the more diversified the chromosomes. However, a larger PS and a higher MR may lead to the requirement of more computational efforts.

To obtain a relative optimal configuration of GA for supporting SCA's variable selection, PS and CF ($\text{MR} = 1 - \text{CF}$ in our GA program) were chosen to examine their effects on fitness. We changed one of them in its possible range and kept the other at its default value during each observation. The results indicated that, when the PS increased from 20 to 110 by every 10 units, the obtained optimal fitness value would fluctuate and keep as a constant after the PS exceeds 100. Similarly, the obtained optimal fitness value would keep the same when CF was greater than 0.8 (in the range from 0 to 1 by every 0.05). Thus, the optimal setting of GA for SCA was set as: population size = 100, crossover fraction = 0.8 and maximum allowable generation = 25. The operation processes of GA under such a setting are illustrated in Fig. 4. Over the entire GA iteration, the best,

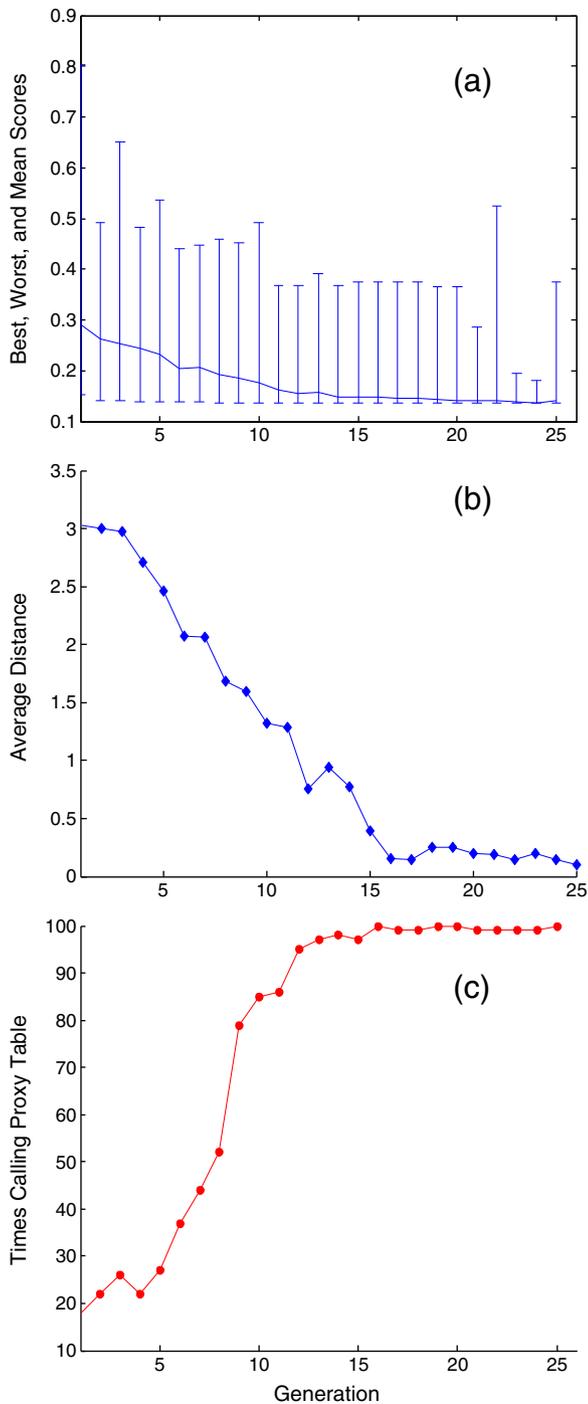


Fig. 4. Representative operation processes of GASCA₂.

worse and the average fitness value (the vertical bars in Fig. 4a) would decrease gradually with the increase of generation number. There were always significant differences between the best and the average fitness values at each generation; the average distance would decrease instantly but automatically keep fluctuating during the process of GA operation (Fig. 4b). This demonstrated that the population diversity could be maintained and the undesired prematurity could be avoided during GA operations.

4.2. Interactions between SCA and GA

Within the GASCA framework, reduction of SCA's computational efforts is critical for improving GASCA's efficiency. This is due to: (i) the

SCA's calculation amount increases significantly with the increase of dataset volume, and (ii) SCA is frequently invoked to calculate the fitness function during the GA operation. Thus, effective use of SCA's results that have been calculated during the former generations would greatly improve the entire efficiency of GASCA. To this end, a proxy table are introduced to realize this mechanism. During GA operations, when GASCA needs to evaluate the fitness of a new chromosome, a proxy table can be used to find out whether the same chromosome has been calculated in the previous generations. Fig. 4c shows that the time to invoke the proxy table gradually increase to its maximum value with the increase of generation numbers. During the early stage of GASCA, the invoking time was relatively limited. It implied that most of the SCAs needed to be trained since there were few chromosome records in the proxy table. At later stages, the invoking time kept increasing since more and more chromosome records were saved in the proxy table. Therefore, the corresponding fitness values of many SCAs could be obtained directly from the proxy table since these chromosomes and their fitness values had already been recorded. In Fig. 4c, the sum of time invoking the proxy table was 1879 while that required for calculating the fitness was 2600; about 72% efforts for fitness calculation had been saved. It thus demonstrated that the introduction of a proxy table could effectively reduce the chance of repetitive training and thus significantly improve the efficiency of GASCA iteration.

In addition to efficiency, the accuracy of the final result of GASCA is also critical. The manner that GA affects the construction of a SCA tree largely determines the performance of a GASCA tree. The factors for controlling SCA's performance consist of dataset quality, dataset partition strategy, SCA's internal parameters (α_{cut} , α_{merge} and N_{min}), and the combination of candidate independent variables (Qin et al., 2007). In this study, the candidate independent variables and inherent parameters of SCA are considered as the major factors that GA would affect a SCA tree. In our practice for the current dataset of food waste composting, it is found that the SCA would tend to be over-fitting (i.e. the RMSE for the test set is much worse than the one for the training set) in one of following settings: 1) when the α_{cut} is less than 0.03 or greater than 0.065; 2) when the α_{merge} is less than 0.05 or greater than 0.065; 3) when the α_{cut} is greater than α_{merge} ; 4) and when N_{min} is less than 7 and greater than 10. Thus, the variation range of inherent parameters are set as: $\alpha_{\text{cut}} \in [0.03, 0.065]$, $\alpha_{\text{merge}} \in [0.05, 0.065]$, and $N_{\text{min}} \in \text{integer in } [7, 10]$. Since α_{cut} and α_{merge} are real numbers, it is difficult to incorporate them into binary-bit chromosomes. Alternatively, a set of representative values for these inherent parameters are chosen as: $\alpha_{\text{cut}} \in \{0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06, 0.065\}$, $\alpha_{\text{merge}} \in \{0.05, 0.055, 0.06, 0.065\}$, and $N_{\text{min}} \in \{7, 8, 9, 10\}$. Meanwhile, 3, 2 and 2 binary-bits are used to represent 2^3 , 2^2 , and 2^2 numbers of representative values, respectively. For example, the last seven bit of chromosome (1010110) includes three parts 101 (No.6 in α_{cut} 's representative values), 01 (No.2 for α_{merge}) and 10 (No.3 for N_{min}), which means the corresponding values were chosen as: $\alpha_{\text{cut}} = 0.055$, $\alpha_{\text{merge}} = 0.055$, and $N_{\text{min}} = 9$. In this manner, the values of chromosomes in GA represent the locations of parameters in the given ranges. Along with the evolution of GA, the performance of the chromosomes will be improved and the related parameters will be transferred to SCA.

4.3. GASCA vs. SCA

To compare GASCA with SCA, four optional SCA/GASCA models were calculated (Table 2). SCA₁ has all of the candidate independent variables and uses each internal parameter at its default value; SCA₂ selects these variables/parameters through the trial and error method; GASCA₁ uses GA to select only independent variables for SCA and employs default SCA's inherent parameters; GASCA₂ uses GA to select both independent variables and SCA's inherent parameters. The results indicated that the R^2 for training sets of GASCA were much higher than those of SCA [R^2 -train: GASCA₂ (0.990) \approx GASCA₁

Table 2
Performance of different SCA/GASCA trees.

Model		SCA ₁	SCA ₂	GASCA ₁	GASCA ₂
Selections	X variables	All Xs	Trial and error	GA-aided	GA-aided
	SCA internal parameters	Default	Trial and error	Default	GA-aided
X		1,2,3,4,5,6,7,8,9,10,11,12	1,3,4,5,6,7,9,10,12	2,4,5,7,10,11	2,4,5,7,10
Y		C/N	C/N	C/N	C/N
α _{cut}		0.05	0.04	0.05	0.04
α _{merge}		0.05	0.06	0.05	0.06
N _{min}		10	10	10	9
Layer		25	21	20	17
Node		165	128	123	100
Training set	R ²	0.985	0.977	0.991	0.990
	RMSE	0.770	0.949	0.582	0.615
Test set	R ²	0.946	0.958	0.946	0.945
	RMSE	1.614	1.454	1.566	1.560
Fitness		0.19647	0.18920	0.14024	0.12851

(0.991)>SCA₁(0.985)>SCA₂(0.977)]; the R² for test sets of GASCA were almost the same as those of SCA₁ while SCA₂ had the highest value [R²-test: SCA₂ (0.958)>GASCA₂ (0.945)≈GASCA₁ (0.946)=SCA₁ (0.946)]; the RMSEs for training sets of GASCA were lower than those of SCA [RMSEs-train: GASCA₁ (0.582)<GASCA₂ (0.615)<SCA₂ (0.949)<SCA₁(0.770)]; the RMSEs for test sets of GASCA were lower than those of SCA₁ while SCA₂ had the lowest value [RMSEs-test: SCA₂ (1.454)<GASCA₂ (1.560)<GASCA₁ (1.566)<SCA₁ (1.614)]. As for the configuration, the number of independent variables of GASCA were much less than those of SCA [GASCA₂ (5)<GASCA₁ (6)<SCA₂ (9)<SCA₁ (12)]; the layer number of GASCA were less than those of SCA [GASCA₂ (17)<GASCA₁ (20)<SCA₂ (21)<SCA₁ (25)]; and the node number of GASCA were much less than those of SCA [GASCA₂ (100)<GASCA₁ (123)<SCA₂ (128)<SCA₁ (165)]. Thus, the fitness of the four models could be ranged in a descending order as: GASCA₂ (0.12851)<GASCA₁ (0.14024)<SCA₂ (0.18920)<SCA₁ (0.19647).

The results indicated that the predictive capability of SCA₂ for the test set was better than GASCA₂ while SCA₂'s other performances were worse. This implied that the trial and error method might improve the SCA's performance to some degree. The comparison between GASCA₂ and SCA₁ was further illustrated in Figs. 5 to 7. It showed that the size of GASCA₂ was much smaller than that of SCA₁ while the performance of GASCA₂ was better. Thus, the GASCA trees have better fitting and predictive ability in performance and higher efficiency in computation than the SCA trees when building nonlinear relationships between the state variables and the C/N ratio during food waste composting. The strength of GASCA is derived from the mutual compensation of GA and SCA. Their integration brings about not only a better GASCA tree but

also intense calculation for searching the optimal SCA tree in GA's evolution environment. SCA only needs to calculate all possible Wilks' Λ values in the corresponding dataset; in comparison, GASCA would have to invoke the SCA function to compare all possible Wilks' Λ values in the GA-selected dataset when GASCA needs to evaluate fitness. The computational effort would inevitably increase but the problem could be mitigated by the use of a proxy table. The final GASCA tree was smaller in size but better in performance, which implied its excellent ability to avoid over-fitting, since the models with bigger size would often over-fit the dataset.

The optimal GASCA₂ tree was finally generated to reflect the relationships between the state variables and the C/N ratio (Table 3; Fig. 5). The optimal combination of input variables include mean temperature (X₂), moisture content (X₄), ash content (X₅), NH₄⁺-N concentration (X₇), and log colony count of mesophilic bacteria (X₁₀); the internal configuration parameters are: α_{cut}=0.04, α_{merge}=0.06 and N_{min}=9. The GASCA tree is a forecasting system flexible to reflect C/N variation during food waste composting based on the relationship between the C/N ratio and state variables. For example, let X₂=47.85, X₄=0.6554, X₅=0.0186, X₇=666.92, and X₁₀=8.4627 as a new sample for the GASCA tree. To predict the C/N ratio, we have: X₇>87.51 for the first branch knot so that the sample enters cluster 3 (Fig. 5); X₅≤0.04, so that it enters cluster 6 and then merges into cluster 8; X₇>273.78, so that it enters cluster 12; X₇>624.22, so that it enters cluster 19; X₂≤28.25, so that it enters cluster 30; X₁₀≤8.49 so that it enters cluster 38 and then merges into cluster 46; X₄>0.63, so that it enters cluster 57 and then finally merges into cluster 68 with a prediction value of 11.11 ± 0.26.

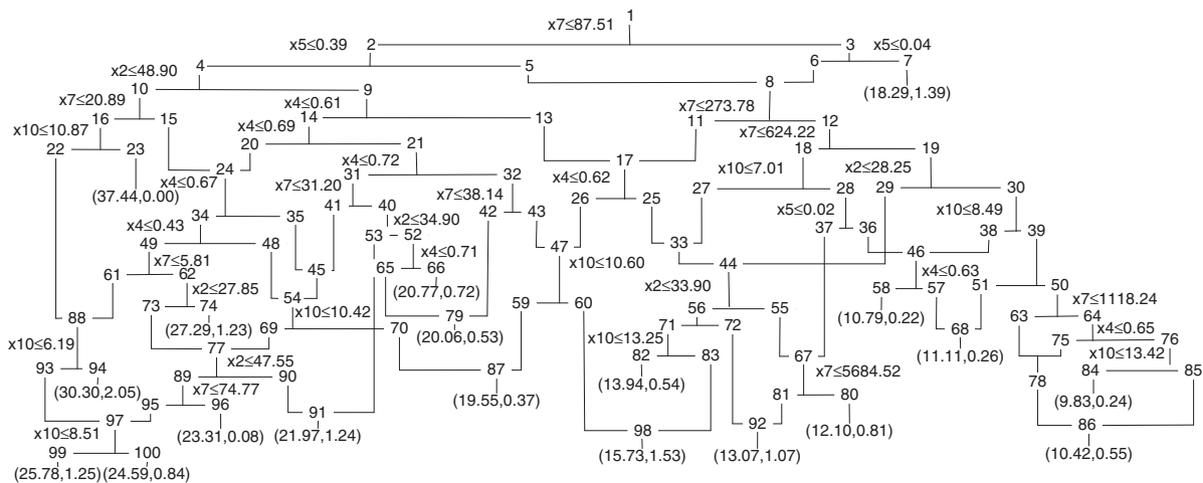


Fig. 5. GASCA₂ tree for the C/N ratio.

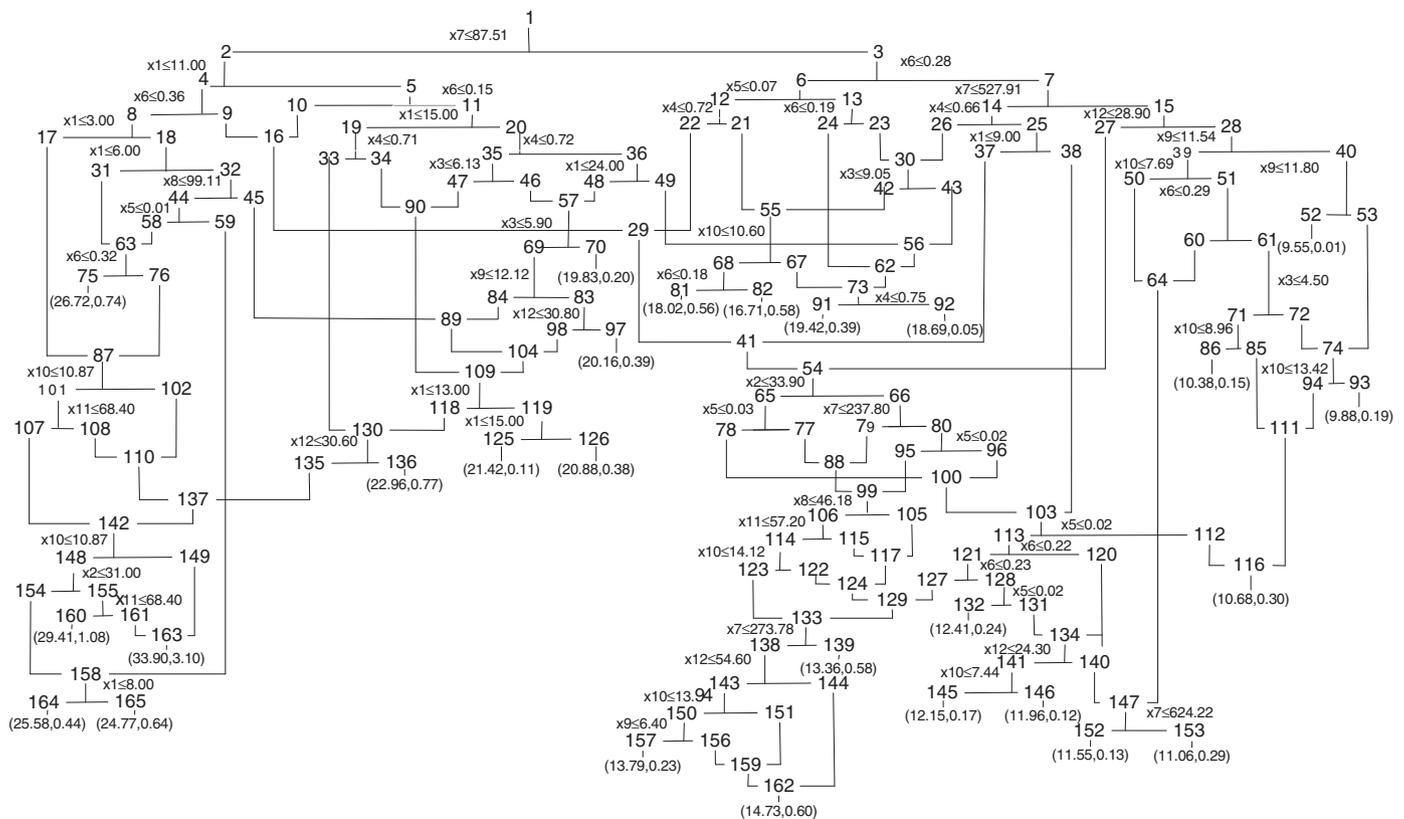


Fig. 6. SCA₁ tree for the C/N ratio.

Based on the configuration of the optimal GASCA tree (Fig. 5 and Table 3), the effect of input (X_k) on the output (Y) in GASCA is quantified using the following equation:

$$Effect(X_k, Y) = \sum_{i=1}^{NN} N_{X_k}(i) \tag{6}$$

where NN is the number of total nodes in the GASCA tree; $N_{X_k}(i)$ is the number of patterns (samples) at node i of the GASCA tree where the corresponding X_k variable is used as the cutting criteria at node i . $N_{X_k}(i)$ represents the number of samples on which X_k variable has effects at node i . According to our previous work (Sun et al., 2009), the effects of the state variables on the C/N ratio could be ranged in a descending order as: $NH_4^+-N > \text{Moisture content} > \text{Ash Content} > \text{Mean Temperature} > \text{Mesophilic bacteria biomass}$. The result might imply that the variation of NH_4^+-N concentration, the associated temperature and moisture conditions, the total loss of both organic matters and available mineral constituents, and mesophilic bacteria activity, might be the important factors affecting C/N ratio during the food waste composting.

4.4. Effect of state variables on the C/N ratio

Dynamic variations of the C/N ratios during the six runs of composting experiments follow different patterns (Fig. 8). In Runs 1 and 2, the initial C/N ratios drop to 14.75 and 15.08 due to the addition of soybean. This is consistent with the conclusion of Wong et al. (2001) where adding soybean residues could lower the C/N ratio because of its relatively high nitrogen content (Wong et al., 2001). The C/N ratios in these two runs gradually drop to the minimum (9.57 and 9.38) in Day 28 and Day 30, respectively, possibly due to the carbon loss by CO₂ emission. The two C/N ratios then gradually rise to 13.70 until Day 37 and to 12.70 until Day 42; then both keep fluctuating or stable, individually. This is possibly due to the decrease of total

nitrogen content caused by NH₃ emission. In Runs 3 and 4, the initial C/N ratios are adjusted to about 32.63 and 29.89. The C/N ratio in Run 3 keeps decreasing to 17.46 until Day 33 and slowly increases to 19.43 until the last day. The C/N ratio in Run 4 keeps decreasing to 19.85 until Day 18, keeps fluctuating to 20.92 until Day 33, and then slowly drops to 17.15 until Day 43. In Run 5, the initial C/N ratio of the composting material is 31.9. The C/N ratio keeps decreasing until Day 35 and become stable (11.79) until Day 39. The decrease of C/N ratio may be associated with the mass loss of organic materials by CO₂ emission (maturing processes). In Runs 6, the initial C/N ratio is raised to 34.99 after adding coal ash amendment. The C/N ratio decreases to 11.88 until Day 15 and fluctuates in a small range between 11.41 and 12.84 in the latter stage. The final composts obtained in all of the six runs are considered mature, since the C/N ratio should be less than 20 for growing plant (Lin, 2008). The good agreement between the predicted and the experimental values of the C/N ratios in Fig. 8 further verifies the good performance of the GASCA tree.

As the base run (without addition of cooked soybean, baking soda and coal ash), Run 5 is chosen to further analyze the relationship between state variables and C/N ratio (Fig. 9). The total composting period in Run 5 lasts for 39 days. A rapid heat-generating process (less than one day) raises the temperature of the composting materials from ambient to thermophilic (higher than 45 °C). This is because of the quick initiation of microbial activities through easy utilization of the high contents of organic materials in the simulated food waste. The mean temperature continues keeping at a high level (higher than 45 °C) for about 25 days and then rising to a peak (52.7 °C) in Day 31. The composting materials then start to cool down and finally reach the ambient temperature (21.15 °C) until Day 39. This is because the heat generation rate becomes less than heat loss rate since the easily degradable materials are consumed gradually. Similar tendencies of temperature could be found in composting processes involving the mixtures of canteen food waste and office waste papers (Bari et al.,

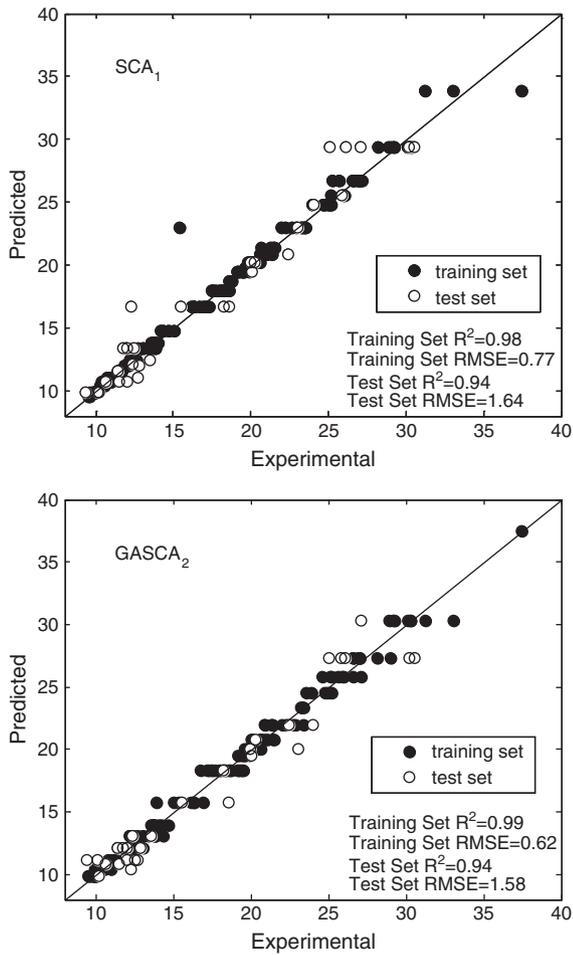


Fig. 7. Prediction comparisons between GASCA₂ and SCA₁ for the C/N ratio.

2000) and a standard mixture of food waste used in Japan (He et al., 2000). From the GASCA tree, the mean temperature is the fourth important variable affecting the C/N ratio.

With the rapid temperature increasing at the initial stage, the thermophilic bacteria start to grow and remain a fluctuant increase during thermophilic period (from 6.08 to 11.94 log CFU/g (dry)). In comparison, the population of mesophilic bacteria decreases gradually from 8.45 to 5.77 log CFU/g (dry) until Day 9, increase gradually in the middle stage of composting (between Day 10 and Day 28), reach a peak (14.20 log CFU/g (dry)) in Day 29, and tend to be stable after Day 32 when the system starts to cool down to the ambient temperature. The mesophilic bacteria are active from Day 10 to Day 28 and are inhibited by the higher temperature stage (during the first several

days and the later stage around Day 30). The activity could be explained by the fact that the mean temperature is in the range only slightly higher than 45 °C (between 44.8 °C and 48.4 °C) during the middle stage (Day 10 to Day 28), which results in that mesophilic bacteria could become more active than thermophilic bacteria. Stagnation or decline in microbial activity in the transition from mesophilic to thermophilic conditions was reported in laboratory-scale reactors when composting food waste or other acid wastes (Schloss and Walker, 2000; Sundberg et al., 2004). This is also reflected in the GASCA tree that mesophilic bacteria biomass is the last important variable affecting the C/N ratio while thermophilic bacteria biomass is excluded.

The pH drops rapidly on the first day and keeps low at the early stage (between 3.99 and 4.89 before Day 27). This is a result of a balance between the generated short-chain organic acids and inhibited microbial activities (Sanchez-Monedero et al., 2001). The pH has a subsequent rapid increase from 4.55 in Day 25 to 7.15 in Day 33, which might be caused by the generation of ammonia by decomposition of nitrogen-containing organic matter and the biodegradation of short-chain organic acids. The pH keeps stable (around 7.12) at the later stage (after Day 33), which is due to reduced production and increased evaporation of the alkaline ammonia (Wong et al., 2001). Similar tendencies of pH could be found in other composting food wastes (Lin, 2008; Chang and Hsu, 2008).

Moisture content is maintained at a level between 62.66% and 75.50%. Its fluctuant increase reflects that water generated by the metabolism of microorganisms is much more than that taken away through aeration. The increase could also be explained by that the decrease rate of compost weight is larger than that of moisture content. The range of moisture contents in this study is considered suitable since many previous studies suggest that 50% moisture content is the minimal requirement for obtaining adequate microbial activities (Liang et al., 2003) and the optimum moisture content for composting could vary significantly for different compost mixtures and stages (Richard et al., 2002). The tendency of moisture content variations is inversely related to the change of the C/N ratio in Run 5 (Figs. 4 and 5), which supports the fact from the GASCA tree that the moisture content is the second important variable affecting the C/N ratio. The reason is that the moisture content could directly affect microbial activities, which further influence the C/N ratio.

The ash content has a fluctuating increase (from 1.48% to 2.80%) while the organic content keeps a decreasing tendency (from 34.25% to 21.92%) during the composting process. This is resulted from both biodegradation of organic matter in the composting materials and mass loss of the composting bulk materials. In Run 5, the ash content has a negative relationship with the change of the C/N ratio, which is reflected in its third important position among all independent variables in the GASCA tree. Since ash content represents the relative mass of conserved mineral constituents (Larney et al., 2005), it

Table 3 Configuration of the optimal GASCA₂ tree.

X	Variables	Number of patterns (locations in the GASCA tree)	Sum of patterns
1	Time	–	–
2	Mean temp (°C)	72(N4), 41(N19), 20(N40), 25(N44), 9(N62), 19(N77)	186
3	pH	–	–
4	Moisture content (%)	58(N9), 55(N14), 22(N17), 33(N21), 27(N24), 11(N34), 10(N46), 14(N52), 18(N64)	248
5	Ash content (%)	77(N2), 90(N3), 11(N28), 26(N39)	204
6	Organic content (%)	–	–
7	NH ₄ ⁺ -N (mg/kg,wet)	167(N1), 76(N8), 14(N10), 57(N12), 22(N31), 11(N32), 10(N49), 23(N50), 16(N56), 19(N67), 16(N89)	431
8	Cumulative NH ₃ emissions (µg/day)	–	–
9	Thermophilic bacteria (log CFU/g (dry))	–	–
10	Mesophilic bacteria (log CFU/g (dry))	9(N16), 16(N18), 35(N30), 9(N47), 19(N54), 9(N71), 15(N76), 9(N88), 14(N97)	135
11	Upper temperature (°C)	–	–
12	Lower temperature (°C)	–	–

Note: M(Ni) represents that there are M patterns in node i of the GASCA tree, where the corresponding X variable is used as cutting criteria.

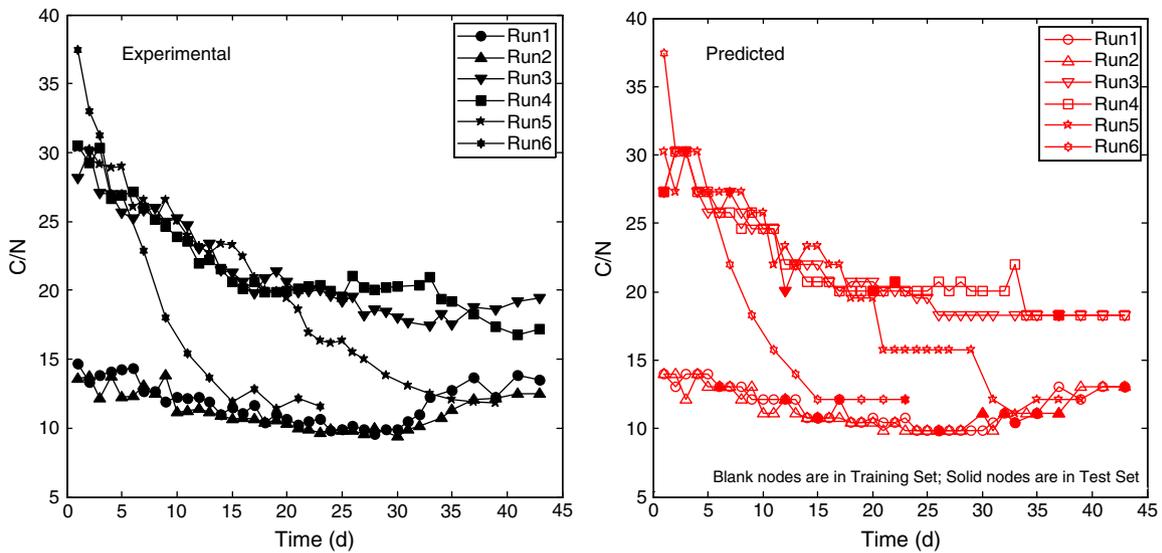


Fig. 8. Dynamic variations of C/N ratios.

implies that the total loss of both organic matters and available mineral constituents have significant effects on the C/N ratio variation.

The $\text{NH}_4^+\text{-N}$ concentration keeps a very low increasing tendency (from 20.86 to 154.41 mg/kg, wet Sample) before Day 21 and rapidly reaches a peak (828.16 mg/kg, wet Sample) in Day 33, which indicates a significant biodegradation of organic nitrogen content (ammonification process) in the latter stage. It is noted that $\text{NH}_4^+\text{-N}$ concentration is the first important variable affecting the C/N ratio in the GASCA tree. This implies that the variation of $\text{NH}_4^+\text{-N}$ concentration plays a major role in the variation of total nitrogen. NH_3 emission in the outlet gas fluctuates within a very small range (between 52.99 and 99.11 $\mu\text{g}/\text{day}$) before Day 29 and increases rapidly to a peak ($5.99 \times 10^4 \mu\text{g}/\text{day}$) in Day 35. The low pH environment in the initial stage (before Day 27) might inhibit the NH_3 release from ammonium. The NH_3 emission peak occurs when

the pH reaches the highest level (7.15) in Day 35. The similar tendencies of NH_3 emission variation and its close relationships with compost pH level were observed in other composting food wastes (Lin, 2008).

5. Discussion

Quantitative analysis of complicated interactions between state variables and the C/N ratio during food waste composting is a challenge. Most previous studies focused on the relationships between initial C/N ratios and final compost quality (Larsen and McCartney, 2000; Eiland et al., 2001; Huang et al., 2004; Kumar et al., 2010) or the effects of initial conditions on the final C/N ratios in compost (Gao et al., 2010). In fact, inherent composting mechanisms exist in dynamic interactions between state variables and target characteristics.

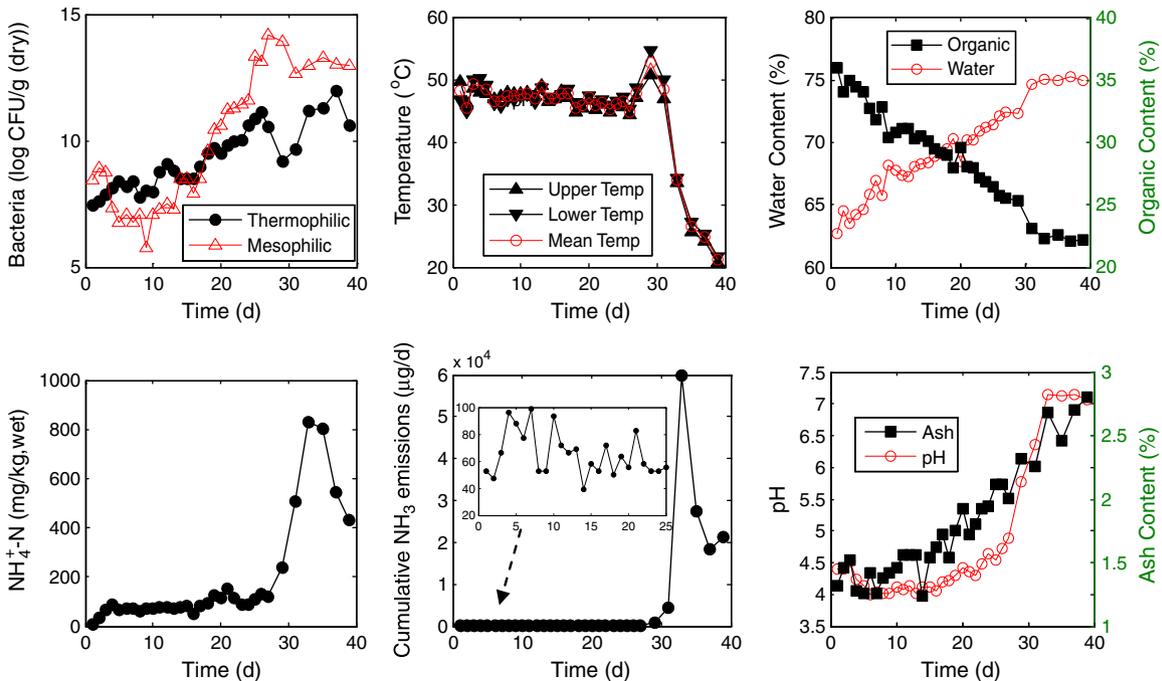


Fig. 9. Dynamic variations of state variables for Run 5.

The statistical models employed to directly quantify these relationships during composting processes are mostly multiple linear regression or response surface analysis (Turner et al., 2005; Khalil et al., 2008; Chang and Chen, 2010). However, the inherently nonlinearity leads to difficulties in applying these conventional models. Simulation-based models can explicitly express dynamic composting mechanisms (Komilis, 2006; Sole-Mauri et al., 2007). However, they are not straightforward to directly interpret such multiple interactions. In comparison, the proposed GASCA model can not only establish a nonlinear relationship between state variables and the C/N ratio but also select more significant state variables and adjust the related statistical parameters automatically.

To ensure a good performance of GASCA, a careful design of fitness function to balance the accuracy and the complexity is indispensable since an optimal SCA largely relies on the expression of fitness function. The convergence efficiency should be further improved through avoiding unnecessary calculation of SCA. A good solution in this study is through introducing the proxy table to store the features of already-calculated SCA trees. Alternatively, more efficient direct-searching algorithms could be introduced in the future to facilitate variable or parameter selection for SCA.

SCA itself still has some weakness besides its high calculation workload during the training phase. Although the effects of the GA-selected state variables on the C/N ratio were quantified in this study, more straightforward methods are required to reflect these effects and to interpret the embedded mechanism. Also, the new samples dropped in the tip-cluster of SCA trees have to possess the same values for the predicted independent variables. If more accuracy of prediction is concerned, other multiple regression models could be developed among samples in the specific tip-cluster to distinguish the sample's difference within the tip-cluster (He et al., 2008a).

In addition, besides the C/N ratio, GASCA can be applied to many other characteristics of interest in waste composting (Chefetz et al., 1996) or co-composting for remediation (Thorn et al., 2002). Especially, some indexes (e.g. mature index, stable index, inactivated pathogen) are expensive or time-consuming for monitoring. Building GASCA-based relationships between the ordinary state variables and these indexes would help identify the most significant relationships, understand the interactive mechanisms, and infer the hard-to-obtain characteristics in an easier manner, during composting and many other environmental processes.

6. Conclusions

Through integrating Genetic-Algorithm (GA) with stepwise cluster analysis (SCA), a GA-aided SCA model (GASCA), was developed to reflect the nonlinear relationships among state variables and the C/N ratio in food waste composting. Six runs of designed experiments through bench-scale reactors in a laboratory were constructed to demonstrate GASCA's performance. The optimal GASCA's inherent parameters were configured as: population size is 100, crossover fraction is 0.8, α_{cut} is 0.055, α_{merge} is 0.055, and N_{min} is 9. A proxy table was introduced to avoid unnecessary and time-consuming calculation of the fitness function, which could save around 70% computational efforts. A GASCA tree with smaller size and better performance than the corresponding SCA tree was obtained. This demonstrated the enhanced ability of GASCA through integration of both GA's variable-screening ability and SCA's nonlinear mapping advantages.

The results showed that GASCA could successfully establish a statistical relationship between the selected state variables and the C/N ratio under discrete and nonlinear complexities. The effects of the GA-selected state variables on the C/N ratio were ranged in a descending order as: NH_4^+ -N concentration > Moisture content > Ash content > Mean Temperature > Mesophilic bacteria biomass. This rank implied that the variation of ammonium nitrogen concentration, the associated temperature and moisture conditions, the total loss of both

organic matters and available mineral constituents, and mesophilic bacteria activity, were the important factors affecting on the C/N ratio during the food waste composting processes. This study, for the first time, combined GA and SCA within a framework to map the relationships in composting processes. It is expected that more direct search algorithms could be coupled with SCA to compensate each other's ability forming more powerful methods to analyze other more complicated relationships during composting.

Acknowledgements

This research was supported by the Major State Basic Research Development Program, Ministry of Science and Technology, People's Republic of China (2005CB724200 and 2006CB403307) and the Natural Science and Engineering Research Council of Canada. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions for improving the manuscript.

References

- An, K., 2006. 'Amendments for ph control in food-waste composting', Master Thesis Faculty of Engineering, University of Regina, Regina, Saskatchewan, 40–59pp.
- Bari QH, Koenig A, Guihe T. Kinetic analysis of forced aeration composting—I. Reaction rates and temperature. Waste Manage Res 2000;18:303–12.
- Cavill R, Keun HC, Holmes E, Lindon JC, Nicholson JK, Ebbs TMD. Genetic algorithms for simultaneous variable and sample selection in metabonomics. Bioinformatics 2009;25:112–8.
- Chang JI, Chen YJ. Effects of bulking agents on food waste composting. Bioresour Technol 2010;101:5917–24.
- Chang JI, Hsu TE. Effects of compositions on food waste composting. Bioresour Technol 2008;99:8068–74.
- Chefetz B, Hatcher PG, Hadar Y, Chen YN. Chemical and biological characterization of organic matter during composting of municipal solid waste. J Environ Qual 1996;25:776–85.
- Chikae M, Ikeda R, Kerman K, Morita Y, Tamiya E. Estimation of maturity of compost from food wastes and agro-residues by multiple regression analysis. Bioresour Technol 2006;97:1979–85.
- Chikae M, Kerman K, Nagatani N, Takamura Y, Tamiya E. An electrochemical on-field sensor system for the detection of compost maturity. Anal Chim Acta 2007;581:364–9.
- Depczynski U, Frost VJ, Molt K. Genetic algorithms applied to the selection of factors in principal component regression. Anal Chim Acta 2000;420:217–27.
- D'Imporzano G, Crivelli F, Adani F. Biological compost stability influences odor molecules production measured by electronic nose during food-waste high-rate composting. Sci Total Environ 2008;402:278–84.
- Eiland F, Klamer M, Lind AM, Leth M, Baath E. Influence of initial C/N ratio on chemical and microbial composition during long term composting of straw. Microb Ecol 2001;41:272–80.
- Foss Analytical AB. Determination of Ammonium in 2 M Potassium Chloride (KCl) Soil Extracts by FIAstar 5000 in FIAstar™ software application note (AN)5226. 2001. Sweden. Ref Type: Pamphlet
- Galvez-Sola L, Moral R, Perez-Murcia MD, Perez-Espinosa A, Bustamante MA, Martinez-Sabater E, et al. The potential of near infrared reflectance spectroscopy (NIRS) for the estimation of agroindustrial compost quality. Sci Total Environ 2010;408:1414–21.
- Gao MC, Li B, Yu A, Liang FY, Yang LJ, Sun YX. The effect of aeration rate on forced-aeration composting of chicken manure and sawdust. Bioresour Technol 2010;101:1899–903.
- Gardner WH. Water content. In: Klute A, editor. Methods of Soil Analysis, Part 1—Physical and Mineralogical Properties. Madison, Wisconsin, USA: American Society of Agronomy and Soil Science Society of America; 1986. p. 493–541.
- Germida JJ. Cultural methods for soil microorganisms. In: Carter MR, editor. Soil sampling and methods of analysis. Canadian Society of Soil Science/Boca Raton: Lewis Publishers; 1993. p. 263–73.
- Giusti E, Marsili-Libelli S. Fuzzy modelling of the composting process. Environ Modell Softw 2010;25:641–7.
- Hamoda MF, bu Qdais HA, Newham J. Evaluation of municipal solid waste composting kinetics. Resour Conserv Recycl 1998;23:209–23.
- He Y, Inamori Y, Mizuochi M, Kong H, Iwami N, Sun T. Measurements of N₂O and CH₄ from the aerated composting of food waste. Sci Total Environ 2000;254:65–74.
- He L, Huang GH, Lu HW. Health-risk-based groundwater remediation system optimization through clusterwise linear regression. Environ Sci Technol 2008a;42:9237–43.
- He L, Huang GH, Lu HW, Zeng GM. Optimization of surfactant-enhanced aquifer remediation for a laboratory BTEX system under parameter uncertainty. Environ Sci Technol 2008b;42:2009–14.
- Herrmann RF, Shann JF. Microbial community changes during the composting of municipal solid waste. Microb Ecol 1997;33:78–85.
- Hsu JH, Lo SL. Chemical and spectroscopic analysis of organic matter transformations during composting of pig manure. Environ Pollut 1999;104:189–96.

- Huang GH. A stepwise cluster-analysis method for predicting air-quality in an urban-environment. *Atmos Environ B* 1992;26:349–57.
- Huang GF, Wong JWC, Wu QT, Nagar BB. Effect of C/N on composting of pig manure with sawdust. *Waste Manage* 2004;24:805–13.
- Khalil A, Domeizel M, Prudent P. Monitoring of green waste composting process based on redox potential. *Bioresour Technol* 2008;99:6037–45.
- Komilis DP. A kinetic analysis of solid waste composting at optimal conditions. *Waste Manage* 2006;26:82–91.
- Kumar M, Ou YL, Lin JG. Co-composting of green waste and food waste at low C/N ratio. *Waste Manage* 2010;30:602–9.
- Larney FJ, Ellert BH, Olson AF. Carbon, ash and organic matter relationships for feedlot manures and composts. *Can J Soil Sci* 2005;85:261–4.
- Larsen KL, McCartney DM. Effect of C:N ratio on microbial activity and N retention: bench-scale study using pulp and paper biosolids. *Compos Sci Util* 2000;8:147–59.
- Liang C, Das KC, McClendon RW. The influence of temperature and moisture contents regimes on the aerobic microbial activity of a biosolids composting blend. *Bioresour Technol* 2003;86:131–7.
- Lin C. A negative-pressure aeration system for composting food wastes. *Bioresour Technol* 2008;99:7651–6.
- Lin YP, Huang GH, Lu HW, He L. A simulation-aided factorial analysis approach for characterizing interactive effects of system factors on composting processes. *Sci Total Environ* 2008;402:268–77.
- Martinez-Sabater E, Bustamante MA, Marhuenda-Egea FC, El-Khattabi M, Moral R, Lorenzo E, et al. Study of the evolution of organic matter during composting of winery and distillery residues by classical and chemometric analysis. *J Agric Food Chem* 2009;57:9613–23.
- Mason IG. Mathematical modelling of the composting process: a review. *Waste Manage* 2006;26:3–21.
- Mudhoo A, Mohee R. Modeling heat loss during self-heating composting based on combined fluid film theory and boundary layer concepts. *J Environ Inf* 2008;11:74–89.
- Nelson DW, Sommers LE. Total carbon, organic carbon, and organic matter. In: Page AL, Miller RH, Keeney DR, editors. *Methods of Soil Analysis, Part 2 Chemical and Microbiological Properties*. Madison, Wisconsin, USA: American Society of Agronomy and Soil Science Society of America; 1982. p. 539–77.
- Qin XS, Huang GH, Chakma A. A stepwise-inference-based optimization system for supporting remediation of petroleum-contaminated sites. *Water Air Soil Pollut* 2007;185:349–68.
- Qin XS, Huang GH, Zeng GM, Chakma A. Simulation-based optimization of dual-phase vacuum extraction to remove nonaqueous phase liquids in subsurface. *Water Resour Res* 2008;44.
- Rao CR. *Advanced statistical methods in biometric research*. New York: Wiley; 1952. 106–207pp.
- Rao CR. *Linear statistical inference and its applications*. New York: Wiley; 1965. 239–301pp.
- Ren YL, Zhang ZY, Ren YQ, Li W, Wang MC, Xu G. Diagnosis of lung cancer based on metal contents in serum and hair using multivariate statistical methods. *Talanta* 1997;44:1823–31.
- Richard TL, Hamelers HVM, Veecken A, Silva T. Moisture relationships in composting processes. *Compos Sci Util* 2002;10:286–302.
- Sanchez-Monedero MA, Roig A, Paredes C, Bernal MP. Nitrogen transformation during organic waste composting by the Rutgers system and its effects on pH. EC and maturity of the composting mixtures. *Bioresour Technol* 2001;78:301–8.
- Schloss PD, Walker LP. Measurement of process performance and variability in inoculated composting reactors using ANOVA and power analysis. *Process Biochem* 2000;35:931–42.
- Sole-Mauri F, Illa J, Magri A, Prenafeta-Boldu FX, Flotats X. An integrated biochemical and physical model for the composting process. *Bioresour Technol* 2007;98:3278–93.
- Sun W, Huang GH, Zeng GM, Qin XS, Sun XL. A stepwise-cluster microbial biomass inference model in food waste composting. *Waste Manage* 2009;29:2956–68.
- Sundberg C, Smøps S, Jsson H. Low pH as an inhibiting factor in the transition from mesophilic to thermophilic phase in composting. *Bioresour Technol* 2004;95:145–50.
- Tatsuoka MM. *Multivariate analysis*. New York: Wiley; 1971. 38–197pp.
- The MathWorks, Inc. *Genetic Algorithm and Direct Search Toolbox(TM) 2 User's Guide*. MathWorks. 2008. Ref Type: Electronic Citation.
- Thomas GW. Soil pH and soil acidity in Sparks. In: D.L., et al, editor. *Methods of soil analysis, Part 3, Chemical methods*. Madison, WI: Soil Science Society of America; 1996. p. 475–90.
- Thorn KA, Pennington JC, Hayes CA. M-15 NMR investigation of the reduction and binding of TNT in an aerobic bench scale reactor simulating windrow composting. *Environ Sci Technol* 2002;36:3797–805.
- Tiquia SM, Tam NFY, Hodgkiss IJ. Microbial activities during composting of spent pig-manure sawdust litter at different moisture contents. *Bioresour Technol* 1996;55:201–6.
- Turner C, Williams A, White R, Tillett R. Inferring pathogen inactivation from the surface temperatures of compost heaps. *Bioresour Technol* 2005;96:521–9.
- Vlyssides A, Mai S, Barampouti EM. An integrated mathematical model for co-composting of agricultural solid wastes with industrial wastewater. *Bioresour Technol* 2009;100:4797–806.
- Wilks SS. *Mathematical statistics*. New York: Wiley; 1962. 20–209pp.
- Wollum II AG. Cultural methods for soil microorganisms. In: Page AL, Miller RH, Keeney DR, editors. *Methods of soil analysis, Part 2 Chemical and microbiological properties*. Madison, Wisconsin, USA: American Society of Agronomy and Soil Science Society of America; 1982. p. 782–95.
- Wong JWC, Mak KF, Chan NW, Lam A, Fang M, Zhou LX, et al. Co-composting of soybean residues and leaves in Hong Kong. *Bioresour Technol* 2001;76:99–106.