The potential of a genetic algorithm and support vector regression to evaluate the organic matter content of manure composts by near infrared spectroscopy

Guangqun Huang and Lujia Han*

Biomass and Bioresource Utilization Laboratory, College of Engineering, China Agricultural University, Beijing 100083, China. E-mail: hanlj@cau.edu.cn

Introduction

China is one of the largest producers of livestock and poultry manure with an annual output of about 3 billion tonnes.[1] The best way to dispose of livestock and poultry manure usually involves aerobic composting.[2] Organic matter content (OMC) is a key factor for the quality evaluation of livestock and poultry manure composts (LPMC). Near infrared spectroscopy (NIRS) is a rapid and cost-effective technique used for the determination of constituents in foods, feed and other commodities.[3] Several studies have shown the potential of NIRS for OMC determination in many kinds of materials.[4–6] In the process of constructing quantitative models using the NIRS technique, the support vector machines (SVM) algorithm has recently been proven as a powerful tool for the analysis of soluble solids content of apple, cimetidine tablets and other materials.[7–11] Genetic algorithms (GA) represent an adaptive heuristic search algorithm that can be successfully applied when the dimension of the data space is too large for an exhaustive search.[12] Tewari[13] and Ying[14] *et al.* successfully explored the estimation of the internal quality in pear and sugars of citrus fruits using a GA algorithm and PLS, with the NIRS technique. In this study, the main goal was to explore the combination performance of a GA and ν-support vector regression (ν-SVR) for OMC determination in LPMC.

Materials and methods

Samples and chemical analysis

A total of 120 LPMC samples were collected from farms and composting factories in 22 provinces of China. Composting materials were cattle manure, chicken manure, pig manure etc. OMC was determined according to the Test Methods for the Examination of Composting and Compost (TMECC) by US Composting Council.[15]

Spectral and statistical analysis

Spectra were recorded using a FT-NIRS system (SPECTRUM ONE NTS; PerkinElmer, New Jersey, USA). Each original sample was scanned three times with 64 co-added scans from 10,000 cm^{-1} to 4000 cm^{-1} in a small cup with a quartz window of 4 cm diameter, and the recorded spectrum which consists of 3001 data points was the average of three times with the format log 1/*R*. All samples were divided into a calibration set (3/4 of the samples) and a validation set (1/4 of the samples). The NIRS calibrations were constructed by means of PLSR and ν-SVR. Varieties of spectral pretreatments were tested: derivative, standard normal variate (SNV) and multiplicative scatter correction (MSC), and others. Statistical parameters were the coefficient of determination in the calibration set (R^2), the standard error of estimation (*SEE*), the coefficient of determination in the validation set (R_v^2), the standard error of prediction (*SEP*) and the ratio of the standard deviation of the reference data in the validation set to the *SEP* (*RPD*).

For SVR, each spectrum has 3001 data points. Different combinations of spectral data points by the GA and 1st derivative pretreatment were used explored as the inputs of ν-SVR using a Radial Basis Function as kernel (ν-RBF-SVR). Three key parameters (ν, C, γ) were optimised for ν-SVR using the step grid search method. SVR was performed using the LIBSVM software package on the platform of MATLAB 7.0.[16]

Results and discussion

Table 1 showed the composition of OMC of total samples in different sets on a fresh weight basis.

Figure 1 shows the NIR spectra of the original samples.

Firstly, PLSR with 1st derivative pretreatment was used to construct a NIRS model. The values of R_v^2, *SEP* and *RPD* were 0.94, 35.07 g kg^{-1} and 4.05, respectively. According to the guidelines,[2] the constructed model had good ability.

Leardi *et al.* found the GA algorithm especially suited to select the best subset for regression for spectral data.[17,18] Several key characteristics used in their work were listed, including (a) response: cross-validated % explained variance; (b) population size: 30 chromosomes; (c) probability of mutation: 1%; (d) number of evaluations per run: 100; (e) window size for smoothing: 3; (f) regression method: PLS.

Table 1. Statistics of OMC in livestock and poultry manure compost samples.

Category	Sample number	Mean	SD	Max.	Min.
Total	120	387.00	142.11	642.42	100.37
Calibration set	90	386.04	142.94	642.42	100.37
Validation set	30	389.91	141.94	624.91	126.49

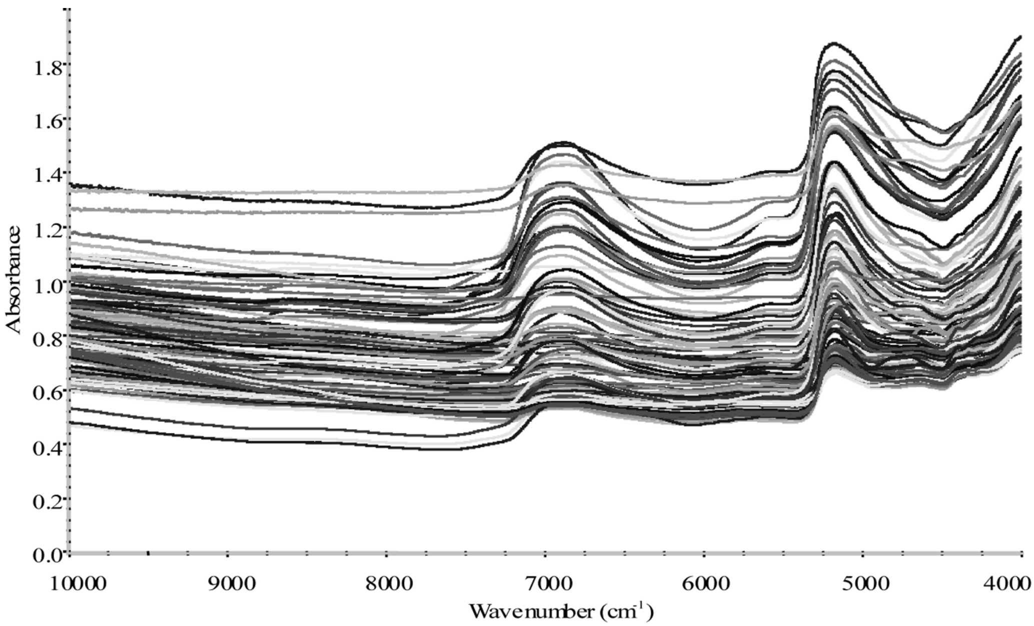

Figure 1. The NIR spectra of livestock and poultry manure composts.

After the GA run, the selected wave numbers were used as the input of ν-RBF-SVR. Table 2 showed the statistics of constructing GA-SVR models.

When the 500 data points selected before 1st derivative pretreatment were used for ν-RBF-SVR, the performance was much better than that by PLSR. But comparatively, 58 data points selected

Table 2. Statistics of ν-SVR models using GA for the wavenumbers selection.

Validation (GA before 1st derivative for SVR)					Validation (GA after 1st derivative for SVR)				
Data points	R_v^2	*SEP*	*RPD*	*Bias*	Data points	R_v^2	*SEP*	*RPD*	*Bias*
58	0.94	33.75	4.21	–0.04	82	0.91	41.67	3.41	6.94
250	0.95	33.35	4.26	2.12	250	0.94	36.07	3.94	4.28
500	0.95	31.86	4.46	1.69	500	0.94	34.91	4.07	1.39
1000	0.94	34.29	4.14	1.95	1000	0.94	33.47	4.24	2.06
2000	0.95	31.90	4.45	1.62	2000	0.94	34.15	4.16	0.55
3001	0.94	34.86	4.07	2.03	3001	0.94	34.86	4.07	2.03

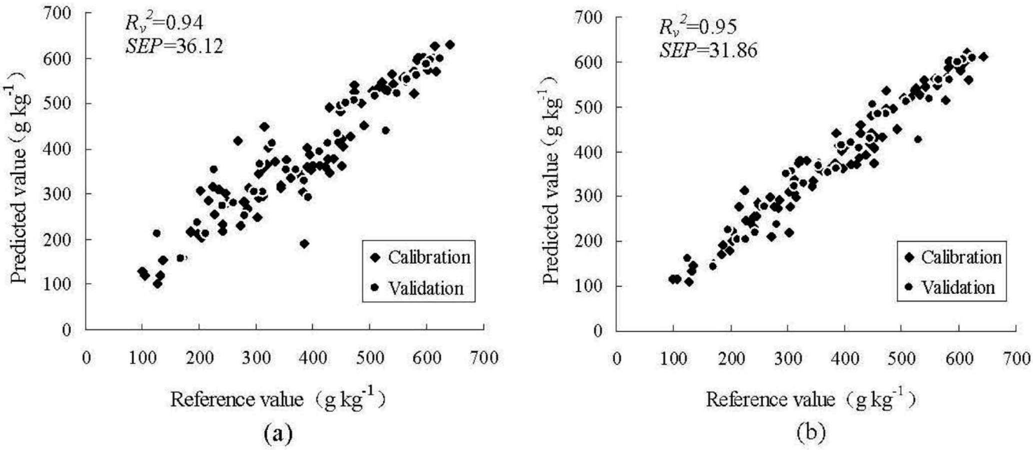

Figure 2. Scatter plots of NIR-predicted vs reference values using NIRS and GA-SVR.

before 1st derivative pretreatment are recommended because of the better performance of prediction with lesser spectral information. When 500 data points and 58 data points selected before 1st derivative pretreatment were used as the input of SVR, the value of ν used for regression was 0.7, and the other two optimal parameters were: $C = 6.86 \times 10^7$, $\gamma = 0.01$ and $C = 3.25 \times 10^6$, $\gamma = 0.25$, respectively. Scatter plots of the best NIR calibrations and validations using PLSR and GA-SVR (500 data points before 1st derivative) are shown in Figure 2, respectively, with serial numbers of “(a)” and “(b)”.

In this study, the results showed that the GA run before mathematical preprocessing may be a good choice. But the optimal combination should be explored for different parameters.

Conclusion

The results showed high efficiency of NIRS for the determination of OMC in livestock and poultry manure composts, using the combination of GA and ν-SVR. In this study, the results showed that the GA run before mathematical preprocessing may be a good choice. But the optimal combination should be explored for different parameters. For higher efficiency, further research is needed to develop SVM-GA method and to optimize key parameters of SVR.

Acknowledgements

This study is supported by the national TCM projects in the 11th Five-Year Period of Ministry of Science and Technology of P.R. China (No. 2006BAD07A13–2 and No. 2006BAD10B05).

References

1. F.S. Zhang, W.Q. Ma and X.P. Cen, *Research and Application of Nutrient Resources Integrated Management Technology*. China Agricultural University Press, Beijing, China (2006).
2. D.F. Malley, C. McClure, P.D. Martin, K. Buckley and W.P. McCaughey, *Commun. Soil Sci. Plant Anal*. **36,** 455 (2005).
3. C. Pasquini and J. Brazil, *J. Braz. Chem. Soc*. **14(2),** 198 (2003).
4. A. Takashi, S. Shimizu, T. Koga and M. Sato, *Jpn. J. Soil Sci. Plant Nutr*. **64,** 669 (1993).
5. M. Nakatani, Y. Harada, K. Kaga and T. Osada, *Jpn. J. Soil Fertil. Soc*. **66,** 159 (1996).
6. J.N. Jae and S.H. Lee, *J. Korea. Chem. Soc*. **44,** 410 (2000).
7. A. Borin, M.F. Ferrao, C. Mello, D.A. Maretto and R.J. Poppi, *Analyt. Chim. Acta*. **579,** 25 (2006).
8. I.V. Kovalenko, G.R. Rippke and C.R. Hurburgh, *J. Agric. Food Chem*. **54,** 3485 (2006).
9. D.Z. Zhu, B.P. Ji, C.Y. Meng, B.L. Shi, Z.H. Tu and Z.S Qing, *Anal. Chim. Acta* **598,** 227 (2007).
10. Y. Zhang, Q. Cong, Y.F. Xie, J.X. Yang and B. Zhao, *Spectrochim Acta Part A: Mol Biomol. Spectrosc*. **71,** 1408 (2008).
11. L.W. Liang, B. Wang, Y. Guo, H. Ni and Y.L. Ren, *Vibr. Spectrosc*. **49,** 274 (2009).
12. A. Durand, O. Devos, C. Ruckebusch and J.P. Huvenne, *Analyt. Chim. Acta*. **595,** 72 (2007).
13. J.C. Tewari, V. Dixit, B.K. Cho and K.A. Malik, *Spectrochim Acta Part A: Mol Biomol. Spectrosc*. **71,** 1119 (2008).
14. Y.B. Ying and Y.D. Liu, J. *Food Eng*. **84,** 206 (2008).
15. US Composting Council, *Test Methods for the Examination of Composting and Compost*. Bethesda, MA, USA (2002).
16. C.C. Chang and C.J. Lin, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm (2001).
17. R. Leardi, "Genetic algorithms in feature selection", in *Genetic Algorithms in Molecular Modeling*, Ed by J. Devillers. Academic Press, London, p. 67 (1996).
18. R. Leardi, *J. Chemometr*. **14,** 643 (2000).