Insect Science (2013) 20, 254–260, DOI 10.1111/j.1744-7917.2012.01519.x

LETTER TO THE EDITOR

Data Processing System (DPS) software with experimental design, statistical analysis and data mining developed for use in entomological research

Qi-Yi Tang and Chuan-Xi Zhang

Institute of Insect Science, Zhejiang University, Hangzhou, China

Abstract A comprehensive but simple-to-use software package called DPS (Data Processing System) has been developed to execute a range of standard numerical analyses and operations used in experimental design, statistics and data mining. This program runs on standard Windows computers. Many of the functions are specific to entomological and other biological research and are not found in standard statistical software. This paper presents applications of DPS to experimental design, statistical analysis and data mining in entomology.

Key words data mining, DPS, entomological research, experimental design, software, statistical analysis

Introduction

A review of the recent agri-biological literature is convincing that quantitative methods in biology have undergone extensive improvements. Nevertheless, many entomologists still hesitate to apply such methods to their data. One reason for this has been the difficulty in acquiring and using appropriate data analysis software (Hammer *et al.*, 2001).

A new version of Data Processing System (DPS) was recently developed to minimize such obstacles and assist researchers and students in agriculture, entomology and other biological fields. The new DPS takes full advantage of the Windows operating system with a modern, spreadsheet-based user interface and extensive graphics. Most DPS algorithms produce graphical output automatically, and the high-quality figures can be printed or pasted into user documents. More than 600 functions are found in DPS, including experimental design, statistical analysis and data mining (see http://www.statforum.com/dps/dps_menu.txt). Many functions in DPS are in a single program package, providing a consistent user interface and minimizing time spent learning a new program. The general linear model (GLM) function in DPS can handle all types of entomological experimental design analyses of variance with a visual interface, such as the mixed multi-factor split-plot design and the lattice design analysis of variance (ANOVA). Some non-statistical analysis functions, such as fuzzy mathematical methods, gray system methods, various types of linear programming, nonlinear programming, analytic hierarchy processes, back propogation (BP) neural networks, radial basis functions (RBF) and data envelopment analyses, can also be found in DPS (Yu *et al.*, 2009). An important aspect of DPS is the tutorial (Tang & Feng, 2007), which includes a large number of data sets to illustrate possible uses of the algorithms. Working through the tutorial allows the user to efficiently obtain a practical overview of the different methodologies.

Some applications of DPS for experimental design, statistical analysis and data mining in the disciplines of entomology and agriculture are presented here. To demonstrate the methods of statistical analysis in DPS, the stable isotopes δ^2H, δ^{18}O, δ^{15}C and δ^{13}N in 39 samples of

Correspondence: Qi-Yi Tang and Chuan-Xi Zhang, Institute of Insect Science, Zhejiang University, Hangzhou 310058, China. Tel/fax: +86 571 88982991; email: qytang@zju.edu.cn, chxzhang@zju.edu.cn

Nilaparvata lugens collected from 13 regions in southern China were analyzed. Several methods, including data description, ANOVA, regression analysis, discrimination analysis, hierarchical clustering and factor analysis, were applied to characterize geographic effects and interpret the relationship among the stable isotopes in *N. lugens*. The original stable isotope data are shown in a table on our website (http://www.statforum.com/dps/table1.xls). To demonstrate the multivariate analysis abilities of DPS, the concentrations of 25 elements (Mg, Al, Ca, Ti, Mn, Fe, Co, Ni, Cu, Zn, Ga, As, Se, Rb, Sr, Zr, Mo, Ag, Cd, Sb, Ba, La, Ce, Nd and Hg) in 78 samples from 2009 in nine regions in southern China were determined, and the data were logarithmically transformed (http://www.statforum.com/dps/elements.xls).

Experimental designs

Scientific planning of the various operations in biology is based on proper experimentation to yield statistically valid and easily verifiable results. The experimental design functions provided in DPS assist biological researchers in designing the appropriate experiments. The experimental designs include a completely randomized design, a randomized complete block design, a Latin squares design, factorial designs, orthogonal designs, split plot designs, augmented designs, uniform designs and mixture designs.

The response of a biological process to various factors is generally nonlinear and has many interactions among those factors. All influential factors must be studied simultaneously in a single experiment. Due to the curvature of the expected response and the presence of interactions among the factors, the size of experiments can grow very large. A class of experimental designs named central composite designs (CCD) were included in DPS; the CCD reduce the number of treatments required to estimate all the terms of a second-order polynomial equation without any loss of efficiency compared with the full factorial design.

The Latin square design using DPS can be found in Xu *et al.* (2011). The uniform design algorithm and the creation of uniform design matrix using DPS refers to Zhu *et al.* (2011).

Data description

Descriptive statistics in DPS include: the minimum, maximum and mean values; the population variance, sample variance, population and sample standard deviations; the median, skewness and kurtosis; and the detection of outliers. DPS additionally includes tests of univariate normality, such as the chi-squared test (χ^2), Shapiro–Wilk tests, D'Agostino's K-squared test, the Jarque–Bera test and the Anderson–Darling test. The Pearson type III distribution fitting procedure is widely accepted and has been used to determine flood flow frequency, rainfall intensity and duration, and the frequency age distribution of HIV/AIDS infection. For associations or bio-community data, several diversity statistics can be computed: the number of taxa, the number of individuals, dominance, the Simpson index, the Shannon index (entropy), Fisher's α and rarefaction (Krebs, 1989).

Data on the basic statistics of stable isotopes using DPS are demonstrated in Table S2 on our website (http://www.statforum.com/dps/table2.xls). The basic statistics include the sample size, mean, variance, standard deviation, median, minimum, maximum, and the Wilk's W and *P*-value of normality tests. The basic statistics of the ratios of δ^2H and δ^{18}O stable isotopes showed that the lower latitudes and shorter distance from the coastline were associated with higher ratios. The ratios of δ^2H were -183.42 ± 54.72, 183.82 ± 14.83 and -187.00 ± 25.08 in Longmen, Hengxian and Qingyuan, respectively, which are sites that were closer to the coastline than others. The ratio of H was -228.72 ± 31.36 in Sandu, which is far from the coastline. The characteristics of oxygen were similar to hydrogen, but carbon and nitrogen did not demonstrate such characteristics. A normality test showed that all four elements fit into a normal distribution. Lastly, the box-plot for 25 elements is shown in Figure 1 on our webpage (http://www.statforum.com/dps/fig1.gif).

Statistical analysis for experiments

DPS provides experimenters with the scientific and statistical procedures needed to maximize the knowledge gained from research data. These procedures include ANOVA on sums of squares for balanced data and GLM approach to analyze any type of experimental design, including unbalanced designs and experiments with missing values. With DPS, we can fit statistical models containing factors whether the data are experimental or observational. ANOVA for a classification factor with more than two levels determines whether the levels of effects are significantly different from each other, but it does not determine which levels differ from which other levels. After performing an ANOVA, DPS can automatically run multiple comparisons of means to produce more detailed information about the differences between the means while controlling the error rates for a multitude of comparisons. Su *et al.* (2011) investigated the differences among the transgenic and control lines for growth and physiological and insect-resistance properties by one-way ANOVA ($\alpha = 0.05$) with Duncan's multiple range test for multiple comparisons using DPS. Data from each ANOVA were evaluated to confirm that the corresponding assumption was satisfied. Applications of ANOVA in DPS have additionally been described by other authors (Su *et al.*, 2008; Lv *et al.*, 2011; Cao *et al.*, 2011; Zhou *et al.*, 2011).

In addition to ANOVA, the likelihood ratio test G-test, χ^2 for comparing binned samples, Mann–Whitney U-test and the Kolmogorov–Smirnov association test (non-parametric), and both Spearman's r and Kendall's τ non-parametric rank-order tests were included in DPS to analyze two or more populations. For associations or biocommunity data, the Dice and Jaccard similarity indices can compare associations limited to absence/presence data. The randomization method for comparing associations is also included. Finally, the program can compute correlation matrices and perform a contingency-table analysis. Su *et al.* (2007) analyzed differences in the frequency of fanning behavior among workers of different patrilines on different days using the G-test and applied G-test to test the significance of differences in patriline frequencies between the fanning workers and the whole colony with DPS.

To evaluate the geographic effects of trace elements among the 13 locations, the *N. lugens* data were analyzed using an ANOVA with a two-stage nested design in DPS. Table 3 online (http://www.statforum.com/dps/table3.xls) lists the results, which include the ANOVA table (F-statistics, df and P-value) by location for each stable isotope and the table of multiple comparisons. The results showed significant differences ($P < 0.01$) among the primary factors in the different locations of origin for the four stable isotopes. The results could then be used for a discriminant analysis of the origin of *N. lugens*.

Regression and curve fitting

Data fitting in DPS includes a range of linear and nonlinear functions. Linear regression can be performed with two different algorithms: the standard (least squares) regression and the robust regression (M-estimation) method. Least-squares regression keeps the x values fixed, and it finds the line that minimizes the squared errors in the y values. Robust regression is a form of regression analysis designed to circumvent some limitations of traditional parametric and non-parametric methods. Robust regression methods are designed to minimize the effects of the violations of assumptions in the underlying data-generating process.

In addition, DPS allows for data fitting of the nonlinear regression equation, such as the logistic equation $y = a/(1{+}be^{-cx})$, Holling's disk equation (Holling, 1959) $N_a = aTN/(1 + aT_hN)$, and the von Bertalanffy growth equation $y = a(1{-}be^{-cx})$, using the Levenberg–Marquardt nonlinear optimization. The equations of nonlinear regression in DPS can also be defined by users.

Count models, such as Poisson regression, logistic regression, probit regression and log-linear models in DPS are a subset of discrete response regression models. Count data are distributed as non-negative integers, are intrinsically heteroskedastic and right-skewed, and have a variance that increases with the mean. Examples of count data include the length of a hospital stay, the number of a certain species of fish in a defined area of the ocean, the number of lights displayed by fireflies over a specified time period, and the classic case of the number of deaths of Prussian soldiers resulting from being kicked by a horse during the Crimean War.

Azzam *et al*. (2010) investigated changes in Cu, Fe, Mn, Zn, Ca, K, Mg and Na contents in rice plants following imidacloprid foliar sprays in the adult female of *N. lugens*. These changes develop from nymphs feeding on treated plants and honeydew produced by females. Multivariate statistical analyses using DPS showed that Fe, Mn and Na in the leaf blades and Fe and Mn in the leaf sheaths could be proportionally transferred to *N. lugens*. The relationship between most elements in adult female bodies and the honeydew showed a positive correlation coefficient. There were significant differences in the contents of some elements in the rice plants and the *N. lugens* from different regions.

Zang *et al.* (2011) investigated host feeding in relation to host density by fitting the Holling's disk equation, $N_a = aTN/(1 + aT_hN)$, where N_a is the number of whiteflies killed by host feeding or parasitism, N is host density, T is the exposure time and T_h is the handling time per host. The data were fitted using DPS. The results showed significant increases in host feeding by host density for mated or unmated whitefly parasitoids ($P < 0.01$) in four cases.

A more complicated non-linear model can be conducted in DPS. Xu *et al.* (2011) conducted a cabbage crop experiment to monitor the population dynamics of pests and native natural enemies and to confirm the effectiveness of relay-intercropping and plant residual mulching in natural enemy conservation and the consequent biological pest control. During the growth period of the crops, the density of Lepidoptera larvae and three kind of predators, namely frog, spider and carbide, were examined and analyzed with a mathematic model as $P_1 = P_{B1} + \{P_{max} - P_{B1} - P_{Y1}[1 - \beta_1\ (t-\tau)^2]\}$ $\mathrm{EXP}[-\alpha_1(t-\tau)^2] + P_{Y1}[1-\beta_1(t-\tau)^2]\ (t <= \tau)$ and $P_2 = P_{B2} + \{P_{max} - P_{B2} - P_{Y2}[1-\beta_2(t-\tau)^2]\})\mathrm{EXP}[-\alpha^2(t-\tau)^2]$

$+ P_{Y2}[1-\beta_2(t-\tau)^2](t >= \tau)$. The model was fitted successfully using DPS.

The theoretical functional relationship of ratios of the stable isotopes δ^2H and δ^{18}O is $\delta^2H = 10 + 8\delta^{18}O$. A total of 39 samples in Table 4 at the DPS website (http://www.statforum.com/dps/table4.xls) yields the regression equation of $\delta^2H = -365.39 + 9.56\delta^{18}O$ ($r = 0.735\,3$, $P<0.000\,1$) with DPS. The regression of delta δ^2H on δ^{18}O showing the individual confidence interval estimates of the mean response is shown in Figure 2 online (http://www.statforum.com/dps/fig2.gif). The intercept of the equation was extremely small, which showed that the ratio of δ^2H in *N. lugens* had little δ^2H related to δ^{18}O. The slope of the equation was close to the theoretical value of 8.

Multivariate analysis

Multivariate analysis can be complicated by including physics-based analysis to calculate the effects of variables for a hierarchical "system-of-systems." Biological data sets, whether based on organism occurrences or morphology, often have high dimensionality. DPS includes a large number of methods (principal components, factor analysis, canonical correlation, discriminant analysis and hierarchical clustering) for multivariate data analysis for entomology.

For the *N. lugens* data, linear discriminant analysis (LDA) was applied to the separation of the *N. lugens* samples by their geographic origin. In our study, the best differentiation among the nine locations was the dependent categorical variable. The concentrations of 25 elements were used as independent variables. The group membership of each sample was already known, so the sample differentiation and data classification were expressed as discriminant functions (http://www.statforum.com/dps/table5.xls). To determine the number of linear discriminant functions to retain, Bartlett's χ^2 test was applied. When the *N. lugens* samples were classified as a function of the place of origin, Bartlett's $\chi^2 = 925.486\,8$, df $= 200$, $P < 0.000\,1$. The discriminant functions were estimated to plot the group subjects in the orthogonal space of the functions. The first two eigenvalues accounted for 53.38% and 19.36% of the total variability, and the sum of these accounted for 72.74% of the total variability. The overall classification success was 98.72% (http://www.statforum.com/dps/fig3.jpg).

Hierarchical clustering routines produce a dendrogram showing how and where data points can be clustered. Clustering is one of the most commonly used methods of multivariate data analysis. Both R-mode clustering (groupings of variables) and Q-mode clustering (samples or associations) can be conducted in DPS. Eight different clustering algorithms, nine different indices for abundance data, and ten indices for presence–absence data were applied to the cluster process. To evaluate the geographic relationship of the stable isotopes to the origin of *N. lugens* among the 13 locations, we conducted a Q-type hierarchical clustering. The dendrogram picture is shown in Figure 4. on our webpage (http://www.statforum.com/dps/fig4.gif). Principal component analysis (PCA) is probably the best known and most widely used dimension-reducing technique for reducing the number of variables while retaining much of the information in the original data set. The PCA routine finds the eigenvalues and eigenvectors of the correlation matrix. The eigenvalues, a measure of the variance of the corresponding eigenvectors, are displayed together with the percentages of variance to account for each of these components. A scatter plot of these data projected onto the principal components is provided, along with the option of including the minimal spanning tree, which is the shortest possible set of connected lines joining all points. This may be used as a visual aid in grouping close points. In an application of PCA in DPS, Zhang *et al.* (2011) described four factors, the fungal fragment number, peak area, the Shannon–Weiner index and the Margale index. The variation in the biomass maximum zone and the complex impacts on the spatial distributions of phytoplankton biomass and production were also evaluated. A PCA of the concentrations of elements generated the first two components, which reach the cumulative variance contribution of 60.86% (http://www.statforum.com/dps/table6.xls). The scatter plot of the first two principal components, the minimal spanning tree and 95% confidence ellipses are shown in Figure 5 online (http://www.statforum.com/dps/fig5.jpg).

Factor analysis is a statistical method to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved, uncorrelated variables called factors. The observed variables are modeled as linear combinations of the potential factors plus "error" terms. Factor analysis originated in psychometrics and is used in behavioral sciences, social sciences, marketing, product management, operations research and other applied sciences that use large quantities of data.

To explain the relationships between the four stable isotopes, we applied factor analysis methods to the *N. lugens* data in DPS. For the stable isotope data, the first two factors reached a high cumulative variance contribution, 77.19%. The two common factors may give a reasonable explanation for the four stable isotopes. The first common factor, which included δ^2H and δ^{18}O, can be treated as the water. The second common factor, including δ^{15}C and δ^{13} N, was atmospheric. These common factors give a reasonable explanation for the interrelationship between the four stable isotopes (http://www.statforum.com/dps/table7.xls).

Correspondence analysis (CA) is an ordination method somewhat similar to PCA but for count or discrete data. Correspondence analysis can compare associations containing counts of taxa or counted taxa across associations in biology. Additionally, a CA is more suitable if species have unimodal responses to the underlying parameters, that is, they favor a certain range of the parameter and become rare at lower and higher values. The axes in CA will normally be interpreted in terms of environmental parameters (e.g., water depth or type of substrate temperature).

Biological assays

A bioassay or biological assay is a procedure to measure the effects of a substance on a living organism and to determine the relationship between the amount (i.e., dose or concentration) of an insecticide administered and the magnitude of response in a living organism. In contrast to the common physical or chemical methods, detailed information on the biological activity of a substance is obtained. The general approach of most bioassays is to perform a dilution assay, which measures the biological responses at several doses. The data become a series of response rates across doses, and the statistical analysis for bioassay can be conducted in DPS. Assays of this sort are often called ''quantal response'' assays. The reported potency is usually the median lethal dose (LD_{50}) or concentration (LC_{50}) predicted to yield responses in 50% of tested animals from a dose–response model. Bioassay also may be used to estimate the median lethal time (LT_{50}) which is the time required to kill 50% of a test population with a fixed dose or concentration of an insecticide.

Often, survival times are not observed more precisely than the interval (for instance, a day) within which the event occurred. This type of survival data is known as grouped or interval-censored data. A discrete analogue of the continuous proportional hazards model (Prentice & Gloeckler, 1978; Allison, 1982) is used to investigate the relationship between these survival times and a set of explanatory variables. The complementary log-log models represent an alternative to a probit analysis for this situation, and DPS supports the complementary log-log models.

The estimates for LT_{50} and LD_{50} for baculoviruses infecting insects using DPS were conducted by Guo *et al.* (2006), Ma *et al.* (2006), Xu *et al.* (2008) and Yang *et al.* (2009). With DPS, the statistical analysis of insecticidal activity and the anti-feedant effect of a new type of biocide, GCSC-BtA, on *Plutella xylostella* can be found in Sengonca *et al.* (2006). Feng and Pu (2005) applied the complementary log-log model to evaluate the interactions of *Beauveria bassiana* and imidacloprid with *N. lugens*.

Non-traditional data analysis

There are many non-traditional data analysis methods in DPS with applications in entomology and biology. Both random forest and support vector machines could be applied to regression and classification. Grey system and fuzzy mathematics could be applied to correlation analysis, forecasting models and synthetic evaluations.

Data on the concentration of elements in the *N. lugens* LDA obtained a classification success of 98.72%. Using random forest clustering in DPS, we not only obtained a classification success of 100% but also measured the importance of each variable (http://www.statforum.com/dps/table8.xls).

Use of DPS' non-traditional data analysis methods, including BP Neural network, support vector machine and random forest, have applications to other specialized disciplines, including neural network model concentrations of N, P and dissolved oxygen in a non-point source polluted river (Chen *et al.*, 2010), radial basis function and support vector regression to design optimization of vehicle roof structures (Pan & Zhu, 2011), BP neural network and time series models for predicting soil salt and water content (Zou *et al.*, 2010), and a geo-statistical approach to mapping soil nutrients (Komnitsas *et al.*, 2010), etc.

Conclusion

DPS is a user-friendly and comprehensive package developed for scientific investigation. DPS provides a window on current and future developments in this rapidly evolving research area. Together with user manual and data examples, the package is an ideal research tool for biological research, especially in the area of entomology. Planned future developments include extended functionality for bioinformation. More information about the development and application of DPS can be found on the web, http://www.statforum.com.

Acknowledgments

The authors gratefully acknowledge the support from National Basic Research Program of China (Grant No. 2010CB126200).

Disclosure

This manuscript and the authors of the manuscript are not involved in any potential conflicts of interest, including financial interests and relationships and affiliations.

References

Allison, P.D. (1982) Discrete-time methods for the analysis of event histories. *Sociological Methodology* 1982 (ed. S. Leinhardt). Jossey-Bass Publishers, San Francisco.

Azzam, S., Yang, F., Wu J.C., Geng, J. and Yang, G.Q. (2010) Imidacloprid-induced transference effect on some elements in rice plants and the brown planthopper *Nilaparvata lugens* (Hemiptera: Delphacidae). *Insect Science*, 18, 289–297.

Cao, Y.Q., Li, H. and Xia, Y.X. (2011) Downregulation of pre-rRNA processing gene Mamrd1 decreases growth, conidiation and virulence in the entomopathogenic fungus *Metarhizium acridum*. *Research in Microbiology*, 162, 729–736.

Chen, D.J., Lu, J. and Shen, Y.N. (2010) Artificial neural network modelling of concentrations of nitrogen, phosphorus and dissolved oxygen in a non-point source polluted river in Zhejiang Province, southeast China. *Hydrological Processes*, 24, 290–299.

Feng, M.G. and Pu, X.Y. (2005) Time-concentration-mortality modeling of the synergistic interaction of *Beauveria bassiana* and imidacloprid against *Nilaparvata lugens*. *Pest Management Science*, 61, 363–370.

Guo, Z.J., Ge, J.Q., Wang, D., Shao, Y.M., Tang, Q.Y. and Zhang, C.X. (2006) Biological comparison of two strains of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus. *BioControl*, 51, 809–820.

Hammer, Ø., Harper, D.A.T. and Ryan, P.D. (2001) PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica*, 4:9.

Holling, C.S. (1959) Some characteristics of simple types of predation and parasitism. *The Canadian Entomologist*, 91, 385–398.

Komnitsas, K., Guo, X.Y. and Li, D.L. (2010) Mapping of soil nutrients in an abandoned Chinese coal mine and waste disposal site. *Minerals Engineering*, 23, 627–635.

Krebs, C.J. (1989) *Ecological Methodology*. Harper & Row, New York.

Lv, G.Y., Lou, Z.H., Chen, S.H., Gu, H. and Shan, L.T. (2011) Pharmacokinetics and tissue distribution of 2,3,5,4’ –tetrahydro xystilbene-2-O-β-D-glucoside from traditional Chinese medicine Polygonum multiflorum following oral administration to rats. *Journal of Ethnopharmacology*, 137, 449–456.

Ma, X.C., Xu, H.J., Tang, M.J., Xiao, Q., Hong, J. and Zhang, C.X. (2006) Morphological, phylogenetic and biological characteristics of *Ectropis obliqua* single-nucleocapsid nucleopolyhedrovirus. *Journal of Microbiology*, 44, 77–82.

Pan, F. and Zhu, P. (2011) Design optimisation of vehicle roof structures: benefits of using multiple surrogates. *International Journal of Crashworthiness*, 16, 85–95.

Prentice, R. and Gloeckler, L. (1978) Regression analysis of grouped survival data with application to breast cancer data. *Biometrics*, 34, 57–67.

Sengonca, C., Liu, B. and Zhu, Y.J. (2006) Insecticidal activity and antifeedant effect of a new type biocide GCSC-BtA against *Plutella xylostella* L. (Lep., Plutellidae). *Journal of Pest Science*, 79, 3–8.

Su, S.K., Albert, S., Zhang, S.W., Maier, S., Chen, S.L., Du, H.H. and Tautz, J. (2007) Non-destructive genotyping and genetic variation of fanning in a honey bee colony RID C-9139–2009. *Journal of Insect Physiology*, 53, 411–417.

Su, S.K., Cai, F., Si, A., Zhang, S.W., Tautz, J. and Chen, S.L. (2008) East learns from West: Asiatic honeybees can understand dance language of European honeybees. *PLoS ONE*, 3, 504–508.

Su, X.H., Chu, Y.G., Li, H., Hou, Y.J., Zhang, B.Y., Huang, Q,J., Hu, Z.M., Huang, R.F. and Tian, Y.C. (2011) Expression of multiple resistance genes enhances tolerance to environmental stressors in transgenic poplar. *PLoS ONE*, 6, e24614.

Tang, Q.Y. and Feng, M.G. (2007) *DPS Data Processing System: Experimental Design, Statistical Analysis, and Data Mining*. Science Press, Beijing, China.

Xu, H.J., Yang, Z.N., Zhao, J.F., Tian, C.H., Ge, J.O., Tang, X.D., Bao, Y.Y. and Zhang, C.X. (2008) *Bombyx mori* nucleopolyhedrovirus ORF56 encodes an occlusion-derived virus protein and is not essential for budded virus production. *Journal of General Virology*, 89, 1212–1219.

Xu, H.L., Qin, F.F., Xu, Q.C., Tan, J.Y. and Liu, G.M. (2011) Applications of xerophytophysiology in plant production–The potato crop improved by partial root zone drying of early season but not whole season. *Scientia Horticulturae*, 129, 528–534.

Xu, Q.C., Fujiyama, S. and Xu, H.L. (2011) Biological pest control by enhancing populations of natural enemies in organic farming systems. *Journal of Food Agriculture & Environment*, 9, 455–463.

Yang, Z.N., Xu, H.J., Thiem, S.M., Xu, Y.P., Ge, J.Q., Tang, X.D., Tian, C.H. and Zhang, C.X. (2009) *Bombyx mori* nucleopolyhedrovirus (BmNPV) *ORF9* is a gene involved in the budded virus production and infectivity. *Journal of General Virology*, 90, 162–169.

Yu, X.L., Yu, W.H., Yi, B. and Wang, X.Y. (2009) Artificial neural network prediction of steric hindrance parameter of polymers. *Chemical Papers*, 63, 432–437.

Zang, L.S., Liu, T.X., Zhang, F., Shi, S.S. and Wan, F.H. (2011) Mating and host density affect host feeding and parasitism in two species of whitefly parasitoids. *Insect Science*, 18, 78–83.

Zhang, Z.Y., Lin, W.X., Yang, Y.H., Chen, H. and Chen, X.J. (2011) Effects of consecutively monocultured *Rehmannia glutinosa* L. on diversity of fungal community in rhizospheric soil. *Agricultural Sciences in China*, 10, 1374–1384.

Zhou, G.M., Zhuang, S.Y., Jiang, P.K., Xu, Q.F., Qin, H., Wong, M.H. and Cao, Z.H. (2011) Soil organic carbon accumulation in intensively managed *Phyllostachys praecox* Stands. *Botanical Review*, 77, 296–303.

Zou, P., Yang, J.S., Fu, J.R., Liu, G.M. and Li, D.S. (2010) Artificial neural network and time series models for predicting soil salt and water content. *Agricultural Water Management*, 97, 2009–2019.

Accepted February 16, 2012