EVALUATION OF CORRELATION METHODS FOR CO-OCCURRENCE NETWORK CONSTRUCTION OF RICE CROP HEALTH SURVEY DATA

J. Sith[A,B], N. P. Castilla[B], C. T. Magculia[C], S. Savary[D], I. B. Pangga[A], and A. H. Sparks[B]
[A] University of the Philippines Los Baños, College, Laguna, Philippines
[B] International Rice Research Institute, Los Baños, Laguna, Philippines
[C] Kobe University, 1-1 Rokkodai-cho, Nada-ku, Kobe-shi, Hyogo-ken, Japan
[D] INRA, UMR1248 AGIR, 24 Chemin de Borde Rouge, Auzeville, CS52627, F-31326 Castanet-Tolosan cedex, France

INTRODUCTION

Detailed on-farm surveys are useful sources of data to help understand yield constraints in farmers' fields. These data also help us understand the interactions and importance of pests and the complex relationships within agroecosystems.

Network analysis is a promising tool frequently used to describe the pairwise relationships of a large number of variables. Network analysis of the co-occurrence patterns of pest and disease incidence could offer new insight into pest management. Since different correlation measures lead to structurally different co-occurrence networks and provide different information, selecting the optimal correlation method is critical.

MATERIALS AND METHODS

Survey data collected from 2010 to 2012 in 420 farmers' fields in irrigated lowland rice-growing areas of five Asian countries (India, Indonesia, Philippines, Thailand and Vietnam) were based on "A survey portfolio to characterize yield-reducing factors in rice" (1).

An incidence matrix was generated to hold the observations in R (2). Pairwise scores between variables were computed using Pearson, Spearman, Kendal and biweight midcorrelation measures. Network models were visualized using the "qgraph" package for R (3).

RESULTS

An evaluation of the four correlation measures on the basis of their weights of edges (correlation coefficients) shows two groups clustered according to hierarchical clustering using Euclidean distance. The Spearman and Kendall correlations were grouped in rank-base methods; the other group, Pearson, and biweight midcorrelation, were non-rank-based correlations.

A Shapiro-Wilk test indicated that the data were not normally distributed with many outliers, thus non-rank-based methods were unsuitable. The Spearman correlation measure was selected because of its robustness to noise, outliers, and ability to accurately predict the interactions in the network versus the Kendall correlation measure.

A weighted co-occurrence network analysis with cutoff on p-value of 0.05, bacterial leaf blight, sheath rot, sheath blight, red stripe, and silver shoot formed a cluster, which indicates that these variables occurred simultaneously (Figure 1).

DISCUSSION

Appropriate correlation measures should closely associate with prior knowledge of the biological system. In this study, Spearman's correlation was found to be the most suitable measure as supported by the results that accurately captured a known relationship between brown planthopper and whitebacked planthopper (4).

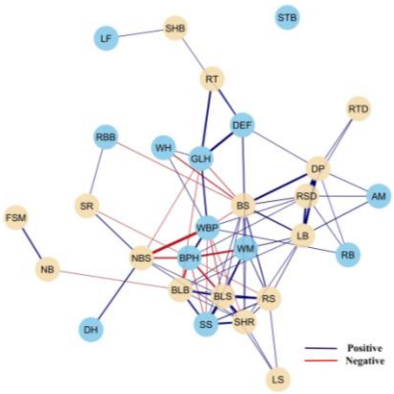

Figure 1. Co-occurrence network of crop health survey data based on Spearman correlation. Blue and brown nodes represent pest injuries and diseases, respectively. The layout of the network graph is based on the Fruchterman-Reingold algorithm, which places nodes with stronger or more connections closer to each other. AM: army worm, BLB: bacterial leaf blight, BLS: bacterial leaf streak, BPH: brown planthopper, BS: brown spot, DEF: defoliator, DH: deadheart, DP: dirty panicle, FSM: false smut, GLH: green leafhopper, LB: leaf blast, LF: leaffolder, LS: leaf scald, NB: neck blast, NBS: narrow brown spot, RB: rice bug, RBB: rice black bug, RS: red stripe, RSD: ragged stunt disease, RT: rat, RTD: rice tungro disease, SHR: sheath rot, SHB: sheath blight, SR: stem rot, SS: silver shoot, STB: stink bug, WBP: whitebacked planthopper, WH: whitehead, WM: whorl maggot.

ACKNOWLEDGEMENTS

Syngenta provided the financial support and collected the on-farm survey data as part of the SKEP collaboration.

REFERENCES

1. Savary, S. and Castilla, N. P. (2009). A survey portfolio to characterize yield-reducing factors in rice. An updated version of IRRI Discussion Paper No 18. 32 p
2. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
3. Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. Journal of Statistical Software, 48(4): 1-18.
4. Win, S. S., Muhamad, R. Ahmad, Z. A. M. and Adam, N. A. (2011). Population Fluctuations of Brown Plant Hopper *Nilaparvata lugens* Stal. And White Backed Plant Hopper *Sogatella furcifera* Horvath on Rice. Journal of Entomology, 8: 183-190.