Ecological Entomology

Ecological Entomology (2017), DOI: 10.1111/een.12401

METHODS

Market basket analysis of grasshopper (Orthoptera: Acrididae) assemblages in eastern Wyoming: a 17-year case study using associative analysis for ecological insights into grasshopper outbreaks

DOUGLAS I. SMITH, MICHAEL F. CURRAN and ALEXANDRE V. LATCHININSKY Department of Ecosystem Science and Management, University of Wyoming, Laramie, Wyoming, U.S.A.

Abstract. 1. This study utilised an associative analysis (AA) technique named market basket analysis (MBA) to investigate whether particular grasshopper (Orthoptera: Acrididae) species associations are common during outbreaks (>9.6 grasshoppers m^{-2}) that last >3 years. This study is the first of its kind to use MBA on animal communities.

2. A subset of the 17 years of grasshopper density data from the Wyoming Grasshopper Survey Dataset was used to explore associations among grasshopper species.

3. Associations of certain species were found with over 80% confidence. Life-history traits of those species commonly found together were examined and compared (*a posteriori*), creating opportunities to hypothesise certain ecological relationships (e.g. interspecific competition, indirect mutualism) for future studies.

4. This case study shows that further MBA analysis of grasshopper assemblages should prove useful in discovering ecological relationships of grasshopper species during outbreaks. Preliminary examples are demonstrated.

Key words. Community ecology, data mining, species association.

Introduction

The Wyoming Grasshopper Survey dataset (WGSD) is a compilation of grasshopper (Orthoptera: Acrididae) and Mormon cricket (Orthoptera: Tettigoniidae) surveys conducted at least annually by several agencies, including Wyoming Weed and Pest Districts and United States Department of Agriculture Animal Plant Health Inspection Service, Plant Protection and Quarantine (USDA APHIS-PPQ). Currently, 214 sites within 14 counties are monitored at least annually. Since 1994, surveys have included identifying grasshopper species from sweep net collections and compiling compositions of grasshopper assemblages. The grasshopper assemblages of the WGSD have yet to be fully analysed. Site-specific grasshopper assemblages can include multiple species, some of which may become predominant during outbreaks (Pfadt, 2002). However, information is scarce on whether one or more species persist over a multi-year outbreak. Grasshopper outbreaks can have dramatic impacts to plant species compositions, habitat structure, carbon cycles and water cycles, especially if several species consume the same type of forage (Latchininsky *et al.,* 2011). Interestingly, the WCDS may comprise as many as 104 different species of Acrididae routinely surveyed at 214 locations, with their densities for nearly 20 years (Lockwood *et al.,* 2015). The only comparable dataset of a single family of animals routinely surveyed over a large geographical area, identified to species and counted (abundance) is that of the yearly Audubon bird survey.

Associative analysis (AA) proves challenging in ecological community research (Gotelli & Ulrich, 2010). Several types of AA (multivariate, Bonferorroni, Bayes M, and Bayes CL) inherently generate type I and type II statistical errors (Gotelli & Ulrich, 2010). In large datasets, AA is cumbersome (computationally) and difficult (several type I and type II errors) to interpret (Tan *et al.,* 2006). In the world of consumer purchasing and marketing, AA is very important in determining which items may be discounted, prompting sales of undiscounted items, placing items near each other to prompt sales of both items, or far from each other to prompt 'impulse purchases' of other products (Surjandari & Seruni, 2005). A very common type of

Correspondence: Douglas I. Smith, Department of Ecosystem Science and Management, University of Wyoming, 1000 East University Ave Laramie, Laramie, WY 82070, U.S.A. E-mail: dsmith59@uwyo.edu

Table 1. Summary of market basket analysis for grasshopper (Orthoptera: Acrididae) species by support [lhs, left-hand sets; rhs, right-hand sets (items yielded from lhs)].

lhs	rhs	Support	Confidence	Lift
Trachyrhachys kiowa	*Ageneotettix deorum*	0.33	0.81	1.3
Amphitornus coloradus	*Ageneotettix deorum*	0.28	0.93	1.5
Melanoplus sanguinipes, *Trachyrhachys kiowa*	*Ageneotettix deorum*	0.23	0.86	1.4
Aulocara elliotti	*Ageneotettix deorum*	0.21	0.89	1.4
Melanolplus bivittatus	*Melanoplus sanguinipes*	0.2	0.86	1.5

Support is the frequency at which species occur together, confidence is the frequency at which the rule is found to be true, and lift is the performance of rule.

AA to investigate purchasing trends within supermarkets, grocery stores, and other commodity stores is the method of market basket analysis (MBA).

Market basket analysis utilises the concepts of Boolean vectors, *a priori* algorithms and association rules to detect relationships between items that are purchased together or placed in the same basket (Agrawal *et al.,* 1993; Agrawal & Srikant, 1994; Tan *et al.,* 2006; Cios *et al.,* 2007; Messaoud *et al.,* 2008; Samecka-Cymerman *et al.,* 2010). Such analysis eliminates bias, type I and type II errors and is completely objective to the investigator. Each transaction consists of Boolean vector sets, the vectors are analysed to determine item sets (i.e. items of different types which are frequently found together), and associative rules state that any item subset is true if the item set is true (Tan *et al.,* 2006). This means that if items i_1, i_2, i_3 and i_4 are 95% likely to be purchased together, then i_1 and i_4 are also 95% likely to be purchased together. The concept and use of MBA is constrained by the number of transactions, not the number of items. A dataset with few transactions may yield differing results than a dataset with multiple transactions, even if the number and type of items are the same. The confidence interval (i.e. the frequency of a rule to be found true), as described by Tan *et al.* (2006), tightens as more transactions of similar item sets increase. MBA uses all elements within the dataset, thus eliminating investigator bias, and the reiterative calculations ignore redundant rules (Tan *et al.,* 2006). In a large database with multiple items where relationships of those items are of interest, MBA can be utilised to find not only common, but also rare or surprising, associations (Samecka-Cymerman *et al.,* 2010; Silva *et al.,* 2016). This method identifies the likelihood of one item, or right-hand sets (rhs) being in a basket if several other items are already in the basket, or left-hand sets (lhs).

The present study uses MBA to find associations of grasshopper species and community structure of grasshopper outbreaks. The MBA method may prove to be a useful tool to answer questions regarding associations of species during grasshopper outbreaks (>9.6 grasshoppers m^{-2}). This study considers portions of the WGSD to be analogous to MBA: customers as WGSD sites, transactions as the annual surveys, and items and item sets as grasshopper species or compositions (respectively) for that year at that site. In an MBA, support is the frequency at which species occur together, confidence is the frequency at which a rule is found to be true, and lift is the measure of performance of an associated rule. This study explores the use of MBA to determine if species associations exist during grasshopper outbreaks (>9.6 grasshoppers m^{-2}) lasting more than 3 years; and if associations exist, they are identified and their biological relationships are examined *a posteriori*.

Methods

A subset (10 of 214 sites) of the WGDS pertaining to grasshopper outbreaks (>9.6 grasshoppers m^{-2}) with the longest duration (3–6 years) was analysed. The MBA used in this study operated the 'arules' package and *a priori* algorithm (support = 0.1, confidence = 0.8, and item set 5.0) of R (R Core Team, 2015). We chose these values so that any result would have >10% support, >80% confidence, and fewer than five item sets (as most grasshopper outbreaks consist of an assemblage of one to five grasshopper species; Pfadt, 2002). Once species associations were identified, we found life-history traits and ecological attributes of those species to compare possible ecological associations.

Results and discussion

Market basket analysis is an exploratory method to find associations within large datasets, and thus any association found is truly without investigator bias. Our MBA indicated that there were 61 items (N), 160 transactions and 36 item sets (or associations) of a possible 3660 associations ($N^2 - N$). The *a priori* rule with the greatest support was lhs{*Trachyrahachys kiowa*} ≥ rhs{*Ageneotettix deorum*}; (support = 0.33, confidence = 0.81). In other words, with 81% confidence, if *T. kiowa* is present then so was *A. deorum* (Table 1). In fact, *A. deorum*, *T. kiowa and M. sanguinipes* were reported in the top five associations for both support and lift (Tables 1). The *a priori* rule with the greatest lift (measure of performance of the associated rule) was lhs{*Melanoplus sanguinipes*, *Phlibostroma quadrimaculatum*} ≥ {*Trachyrachis kiowa*} (Table 2). The support for this rule was 0.1, while confidence was 0.89. This report shows, with nearly 90% confidence, that if a grasshopper assemblage includes *M. sanguinipes* and *P. quadrimaculatum*, it will also include *T. kiowa* (Table 2).

Associating traits (distribution, food preferences and timing of hatching/development) of particular grasshopper species that were frequently surveyed together (item sets) after MBA analysis provides an opportunity to hypothesise ecological relationships without bias. As in Table 2, we first found associations of species than examined their ecological traits. Of the six most

Table 2. Summary of market basket analysis for grasshopper (Orthoptera: Acrididae) species by lift [lhs, left-hand sets; rhs, right-hand sets (items yielded from lhs)].

lhs	Yield	rhs	Support	Confidence	Lift
Melanoplus sanguinipes (HP, EH), *Phlibostroma quadrimaculatum* (SP, IH)	»	*Trachyrhachys kiowa* (SP, IH)	0.1	0.89	2.2
Ageneotettix deorum (MP, EH), *Melanoplus infantilis* (HP, IH), *Melanolus sanguinipes* (HP, EH)	»	*Trachyrhachys kiowa* (SP, IH)	0.14	0.88	2.1
Amphitonus coloradus (MP, EH), *Melanoplus infantilis* (HP, IH)	»	*Trachyrhachys kiowa* SP, IH)	0.11	0.85	2.1
Ageneotettix deorum (MP, EH) *Amphitornus coloradus* (MP, EH), *Melanoplus infantilis* (HP, IH)	»	*Trachyrhachys kiowa* (SP, IH)	0.1	0.85	2.1
Ageneotettix deorum (MP, EH), *Melanoplus infantilis* (HP, IH)	»	*Trachyrhachys kiowa* (SP, IH)	0.17	0.84	2.1

Support is the frequency at which species occur together, confidence is the frequency at which the rule is found to be true, and lift is the performance of the rule with biological traits of grasshopper species found *a posteriori* (adapted from Pfadt, 2002).
Biological traits are abbreviated as follows: HP, highly polyphagous feeding on grasses and forbs; MP, moderately polyphagous feeding mostly on grasses and some forbs; SP, slightly polyphagous feeding mainly on warm-season grasses; EH, early hatching from April to May; IH, intermediate hatching from June to July.

common species of grasshoppers found during outbreaks lasting > 3 years, two are considered highly polyphagous (feeding on a wide variety of grasses and forbs), two are considered moderately polyphagous (feeding mostly on grasses) and two are considered slightly polyphagous (feeding mainly on warm season grasses) (Pfadt, 2002). Three of the six species are considered as early-hatching (April–May), while the other three species are considered as intermediate-hatching (May–June) (Table 2).

These results suggest that MBA is a useful tool for finding associations of species without bias or underlying assumptions. This objective method can lead to inquiries into the ecological explanation of the associations, without a predetermined hypothesis for the association. For instance, interspecific and intraspecific competition (through similar forage preferences of differing species and within species), indirect mutualism (through compensatory forage growth) (Holecheck *et al.*, 1998; Belovsky & Slade, 2000; Branson, 2008) and predator/parasitoid load sharing (Heimpel *et al.,* 2003) may be investigated after determining specific associations. Determining relationships among grasshopper species may yield insights into the population regulatory mechanisms associated with outbreaks in terms of duration, frequency and intensity. These insights could allow for a paradigm shift from treating grasshopper outbreaks as defined by a generic entity ('grasshoppers') to a species-specific process, which includes certain associations of pest species.

Market basket analysis has limitations. For example, counts of grasshoppers are not included in MBA. The population density of one particular species of grasshopper may have an indirect or direct effect on the density of another, although both are present in the survey. Another limitation of MBA concerns the number of transactions and possible item selection. If either the number of transactions or the number of possible items obtained is low, the results could be non-significant (as is the case with any statistical test). Nonetheless, this study reports the first use of MBA to analyse the species association of animals. Future enquires using the MBA technique with the entire WCDS database are likely to provide important insights into species-related processes of grasshopper population dynamics and therefore generate hypotheses regarding ecological factors governing the associations. From a pest management perspective, such a technique could be instrumental in understanding grasshopper outbreak genesis and structure. From a community ecology perspective, as ecological databases become more numerous and detailed, this data-mining technique could become the method of choice to quickly, efficiently and objectively discover species associations *a priori*.

Acknowledgements

We thank the reviewers for their comments and suggestions. This project was funded by the University of Wyoming and does not conflict with other interests.

References

Agrawal, R. & Srikant, R. (1994) Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference*, September, 1994, Santiago, Chile, pp. 487–499.

Agrawal, R., Imielinski, T. & Swami, A. (1993) Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD Conference*, May, 1993, Washington, District of Columbia, Vol. 22, pp. 207–216.

Branson, D. (2008) Influence of a large summer precipitation event on food limitation and grasshopper population dynamics in a northern Great Plains grassland. *Environmental Entomology*, **37**, 686–695.

Cios, K., Pedrycs, W., Swinarski, R. & Kurgan, l. (2007) *Data Mining: A Knowledge Discovery Approach*. Springer, New York, New York.

Gotelli, N. & Ulrich, W. (2010) The empirical Bayes approach as a tool to identify non-random species associations. *Oecologia*, **162**, 463–477.

Heimpel, G.E., Neuhauser, C. & Hoogendoorn, M. (2003) Effects of parasitoid fecundity and host resistance on indirect interactions among hosts sharing a parasitoid. *Ecology Letters*, **6**, 556–566.

Holecheck, J., Pieper, R. & Herbal, C. (1998) *Range management: Principles and practices*. 3rd edn, Printice Hall, Englewood Cliffs, New Jersey, pp. 356.

Latchininsky, A., Sword, G., Sergeev, M., Cigliano, M. & Lecoq, M. (2011) Locusts and grasshoppers: behavior, ecology, and biogeography. *Psyche*, **2011**, 1–4.

Lockwood, J.A., McNary, T.J., Larsen, J.C., Zimmerman, K., Shambaugh, B., Latchininsky, A. *et al.* (2015) *Distribution Atlas for Grasshopper and Mormon Crickets in Wyoming 1987-2016*. University of Wyoming and USDA APHIS PPQ, Cheyenne, Wyoming [WWW document]. URL http://www.uwyo.edu/capsweb [accessed on March 2015].

Messaoud, R.B., Rabaseda, S.L., Missaoui, R. & Bossaid, O. (2008) LOEMAR. Online environment for mining association rules in multidimensional data. *Data Mining and Knowledge Discovery Technologies* (ed. by D. Taniar), pp. 1–35. IGI Global, Hershey, Pennsylvania.

Pfadt, R. (2002) *Western Grasshoppers*. Bulletin 912. Wyoming Agricultural Experiment Stations, Laramie, Wyoming.

R Core Team (2015) *R: Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria [WWW document]. URL https://www.R-project.org/ [accessed on August 2015].

Samecka-Cymerman, A., Stankiewicz, A., Kolon, K., Kempers, A., Rob, S. & Leuven, W. (2010) Market basket analysis: a new tool in ecology to describe chemical relations in the environment-a case study of the fren Athyrium distentifolium in the Tatra National park in Poland. *Journal of Chemical Ecology*, **36**, 1029–1034.

Silva, L., Siqueira, M., Pinto, F., Barros, F., Zimbrão, G. & Souza, J. (2016) Applying data mining techniques for spatial distribution analysis of plant species co-occurrences. *Expert Systems with Applications*, **43**, 250–260.

Surjandari, I. & Seruni, C. (2005) Design of product placement layout in retail shop using market basket analysis. *Journal of Technology*, **9**, 43–47.

Tan, P., Steinbach, M. & Kumar, V. (2006) *Introduction to Data Mining*. Pearson Education Inc., Boston, Massachusetts.

Accepted 21 February 2017

Associate Editor: Chris Hassall