Ecology, 91(7), 2010, pp. 1892–1899
© 2010 by the Ecological Society of America

Reports

Revealing ecological networks using Bayesian network inference algorithms

Isobel Milns,[1] Colin M. Beale,[2,3,4,5] and V. Anne Smith[1,6]

[1]*School of Biology, University of St Andrews, St Andrews, Fife KY16 9TH United Kingdom*
[2]*The Macaulay Institute, Cragiebuckler, Aberdeen AB15 8QH United Kingdom*
[3]*Institute of Biological and Environmental Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen AB24 2TZ United Kingdom*
[4]*Centre for Biodiversity Research, 6270 University Boulevard, University of British Columbia, Vancouver V6T 1Z4 Canada*
[5]*Department of Biology (Area 18), P.O. Box 373, University of York, York YO10 5YW United Kingdom*

Abstract. Understanding functional relationships within ecological networks can help reveal keys to ecosystem stability or fragility. Revealing these relationships is complicated by the difficulties of isolating variables or performing experimental manipulations within a natural ecosystem, and thus inferences are often made by matching models to observational data. Such models, however, require assumptions—or detailed measurements—of parameters such as birth and death rate, encounter frequency, territorial exclusion, and predation success. Here, we evaluate the use of a Bayesian network inference algorithm, which can reveal ecological networks based upon species and habitat abundance alone. We test the algorithm's performance and applicability on observational data of avian communities and habitat in the Peak District National Park, United Kingdom. The resulting networks correctly reveal known relationships among habitat types and known interspecific relationships. In addition, the networks produced novel insights into ecosystem structure and identified key species with high connectivity. Thus, Bayesian networks show potential for becoming a valuable tool in ecosystem analysis.

Key words: avian community; Bayesian network; ecological network; Peak District National Park, UK; species distribution.

Introduction

The complexity of real ecosystems is staggering (Hutchinson 1959): over 1000 species of beetle alone live on a single Panamanian tree species, all of which may interact in many ways, from competition and predation, to facilitation and mutualism (Erwin 1982). For such complex systems, untangling interactions and relationships between species may seem a lost cause. However, in our changing world, understanding ecosystem stability and fragility is of growing importance—yet to do so we must understand the networks that form the systems (Dunne et al. 2002, Montoya and Sole 2002). If, for example, we want to understand how species will respond to climate change, we need to improve the current methods of identifying climate associations to incorporate functional relationships between species (Davis et al. 1998, Pearson and Dawson 2003). Similarly, to understand how a biological control agent may have non-target effects, or knock-on effects of chemical control methods on populations of beneficial insects, we need to identify interactions that exist within the ecosystem (Cohen et al. 1994, Henneman and Memmott 2001). However, revealing networks of relationships within an ecosystem is complicated (Proulx et al. 2005).

To discover information about complex ecological systems efficiently, tools for inferring structure of networks from field data are needed. Previously in ecology this has either meant extraordinarily detailed fieldwork, such as total counts of parasitism events or behavioral observations of competition/facilitation (Memmott et al. 2000*a*, *b*, Proulx et al. 2005); identification of simple relationships between species and habitats through use of classical statistical methods (Guisan and Thuiller 2005, Knight and Beale 2005, Osborne et al. 2007), or more sophisticated modeling of small systems with limited diversity (Krivtsov 2004, Moya-Larano and Wise 2007). Recently, an approach has arisen in biology that is capable of inferring network structures, capturing nonlinear, stochastic, and arbitrary combinatoric relationships: discrete Bayesian networks (BNs; Heckerman et al. 1995). These "Bayesian network inference algorithms" have been applied to reveal gene regulatory networks using gene microarray data (Friedman 2004). Most recently BNs were transferred to a new biological system and shown to reveal known pathways of neural information flow networks from brain electrophysiology data, out-performing previous regression-based methods (Smith et al. 2006). Such a flexible tool capable of identifying the complex relationships

Manuscript received 29 April 2009; revised 30 November 2009; accepted 25 January 2010; final version received 23 February 2010. Corresponding Editor: S. J. Schreiber.
[6] Corresponding author.
E-mail: anne.smith@st-andrews.ac.uk

involved in gene and neuronal regulation potentially offers a valuable new method for recovery of ecological networks.

Here, we evaluate the potential usefulness of BN algorithms for ecological data: we apply BN inference to avian count and habitat data collected from the Peak District National Park, UK. We examine properties of the revealed networks and evaluate them against known features of this ecosystem. In particular, within the Peak District population, we expect many relationships of upland birds with habitat, especially altitude; we expect strong links between Red Grouse and common heather (Pearce-Higgins and Grant 2006); and we expect raptor species to be related positively with their prey species (Sih 2005).

Methods

Data sets.—We used bird observation data, available at a 100-m scale, from surveys of the Peak District National Park (Appendix A: Fig. A1) in 1990 and 2004 (Stillman and Brown 1994, Carr and Middleton 2004, Pearce-Higgins et al. 2006). We excluded birds migrating over the area and birds with single records, leaving 37 upland bird species. Eight species were present in 2004 but not 1990. We additionally used satellite derived environmental variables, also available at 100-m scale: three geographical variables (altitude [m], slope [degrees], and path distribution) and six vegetative variables (vegetation height plus percent cover of burned and unburned heather [*Calluna vulgaris*], grass [*Molinia*, *Nardus*, and fine-leaved grass species], cotton grass [*Eriophorum* spp.], and non-heather dwarf shrub [*Vaccinium myrtillus* and *Empetrum nigrum*; Pearce-Higgins et al. 2006]). See Appendix A for further details of the data used. The 100-m^2 resolution of both bird counts and habitat data were aggregated to create data sets on four spatial scales: 0.5 km, 1 km, 2 km, and 5 km.

Bayesian network inference algorithm.—The BN analysis was performed using Banjo v1.0.5, a publicly available BN algorithm (*available online*)[7]. We refer the reader to Appendix B for a general overview of BNs and Banjo. Briefly, a BN represents statistical dependence between variables (here, species and habitats) by links in the network. BNs preferentially show direct links, not indirect links (i.e., mediated through other measured variables; Heckerman et al. 1995). Banjo uses heuristic search to identify high scoring networks based on a BSM (Bayesian scoring metric) and calculates an influence score for each link, representing the sign (+ or –) and magnitude of influence, with the value 0.0 reserved for non-monotonic relationships (e.g., U-shaped or combinatoric; Yu et al. 2004).

In order to adapt BNs for ecological data, we made several modifications in data discretization, pre-filtering of relationships, and model averaging techniques, details of which are available in Appendix A. Note that although functional relationships between species can be two-way, we present undirected links in our results, because directionality in BNs can be statistically equivalent (see discussion of BN equivalence classes in Appendix B). These links represent the ability to predict one variable from knowledge of another; this is prediction in an informative, rather than causal, sense. For example, equivalent BN representations may show "altitude → Wren" (knowing altitude helps predict presence of wrens) or an equally valid "Wren → altitude" (knowing presence of wrens helps predict altitude); we would report an undirected "altitude – Wren" relationship indicating mutual predictive value.

Our model averaging technique provides information on the probability and average influence score of each link in the network. We defined *highly probable functional relationships* as those in which we have the greatest confidence of being in the network (further details in Appendix A). When searching on the bird plus habitat data sets, connections among habitats were excluded to enable maximum discovery of bird–habitat relationships.

Identification of highly connected species.—We calculated each species or habitat variable's connectivity by measuring "importance": the sum of the probabilities of all links with which a variable was involved. We defined the spatial scale at which individual species were most strongly connected as the scale (0.5, 1, 2, or 5 km) within each year at which it had the highest rank-order importance.

Comparisons across data sets.—To understand how specific the discovered network structures were to their data set, we compared fit of networks across spatial scales for each year and across years for each spatial scale. We took each data set (referred to as the *original data set*) and compared the BSM scores of the top 100 networks (see Appendix A) discovered from that data set (referred to as *original networks*) to those of the networks discovered on other data sets but fitted now to the original data set (referred to as *comparison networks*). This provides a measure of how well a structure learned on a different data set explains the current data: the closer in score the comparison networks are to the original networks, the better the other structures explain the dependencies present in the original data set, and thus the more similar the dependencies in the two data sets.

Comparison with lasso regression networks.—In order to evaluate the difference between BNs and a network building technique more common in ecology, we created networks from the 1-km habitat data—both raw data and the same discrete data presented to the BN—using a lasso regression analysis (an algorithm for linear regression shrinkage and model selection [Tibshirani 1996]). We calculated a matrix of standardized regression coefficients for each variable regressed on all others (number of lasso steps identified by 10-fold cross-

[7] ⟨www.cs.duke.edu/~amink/software/banjo/⟩

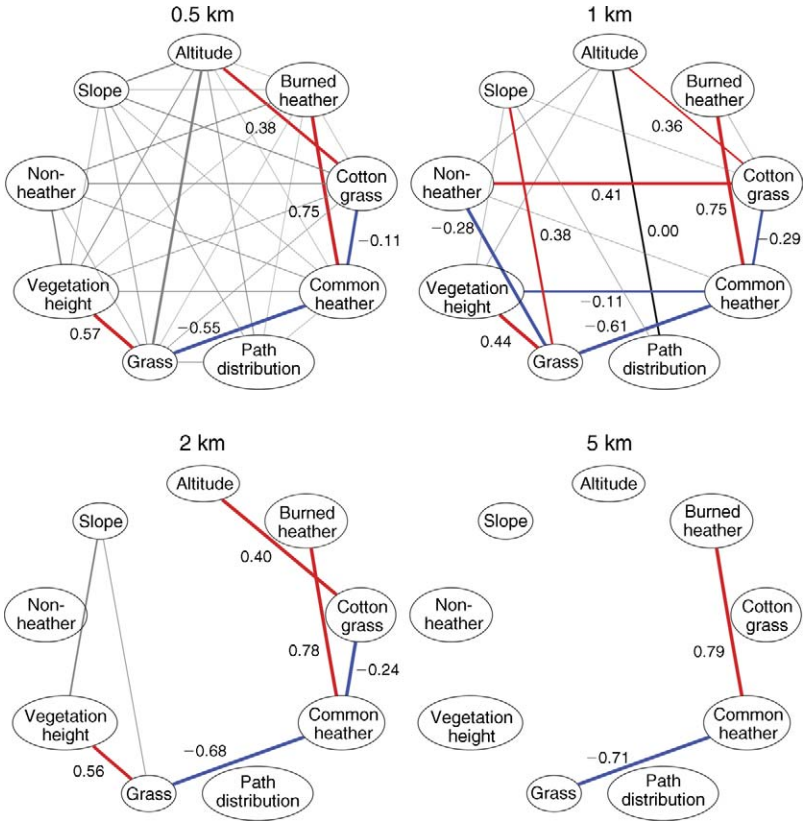


FIG. 1. Inter-habitat networks. Networks for each of the four spatial scales are shown. Habitat variables are enclosed in ovals; lines connecting them represent recovered relationships. Line thickness is scaled to the square of relationship probability. Relationships identified as being highly probable functional relationships are colored red for positive, blue for negative, and black for non-monotonic, and are labeled with mean influence scores; all other relationships are in gray.

validation) and report links with weights representing the average of this matrix folded at the diagonal.

RESULTS

Networks recovered by Bayesian network inference algorithm.—Networks of habitat variables alone are shown in Fig. 1. The greatest number of highly probable relationships were present at 1 km; the highly probable relationships in the 0.5-km and 2-km networks were a subset of these, and those in the 5-km network were a further subset. The influence scores of relationships appearing at three or four spatial scales were similar across these scales, always the same sign (+ or −) and differing on average by less than 0.1 (0.07 ± 0.01).

Networks of habitat plus birds are shown in Fig. 2. Influence scores of relationships were highly skewed toward low values with just a few strong relationships (Fig. 3), and of the highly probable relationships between bird species, most (109 of 120) were positive. Bird–habitat relationships accounted for 35% of highly probable relationships (including most negative relationships). Specific links were found more often at a single spatial scale than expected by chance ($\chi^2 = 20.2$, df $= 1$, $P < 0.001$ for 1990; $\chi^2 = 48.3$, df $= 1$, $P < 0.001$ for 2004). For links present at multiple scales, influence scores were consistent within years, having a mean difference of 0.12 ± 0.01 for 1990 and 0.16 ± 0.02 for 2004. Only one pair of influence scores differed in sign across spatial scales, Dunlin with non-heather in 2004 (−0.19 at 0.5 km and 0.35 at 2 km).

Thirty-one highly probable relationships were found in both years (at any spatial scale). Of these, only one differed in influence score sign across years (Curlew with Snipe, −0.03 in 1990 and 0.16 in 2004). Twenty of these relationships occurred at multiple spatial scales in at least one of the surveys, a much higher proportion than either survey alone ($\chi^2 = 12.2$, df $= 1$, $P = 0.0005$ vs. 1990; $\chi^2 = 24.16$, df $= 1$, $P < 0.0001$ vs. 2004). Nine of the 31 relationships were between birds and habitat, a similar proportion to that found overall ($\chi^2 = 0.28$, df $= 1$, $P = 0.60$).

Variable importance across networks.—Within survey years, variable importance was broadly similar across spatial scales (Kendall's $W = 0.64$ for 1990, 0.56 for 2004, $P < 0.0001$ for both [Legendre 2008; see Appendix A]). All four scales were individually significantly concordant with the others ($P = 0.001$ to $P < 0.0001$), although 0.5 km and 5 km had lower concordance than

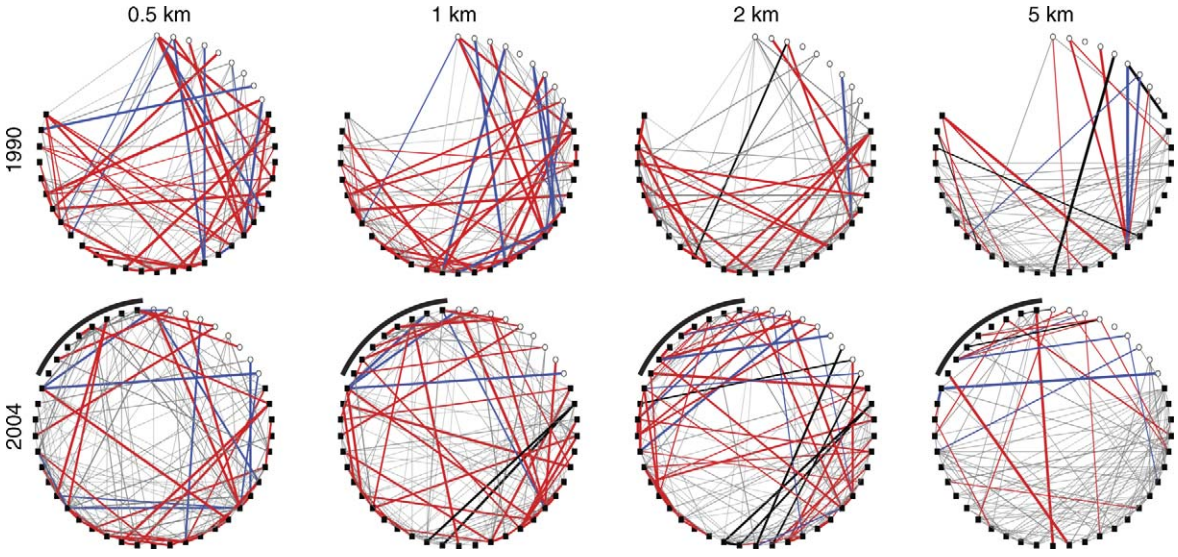


FIG. 2. Interspecific and species–habitat networks. Networks for each sampling year at each of the four spatial scales are shown (details of all data sets with nodes identified are in Appendix C: Figs. C1–C8). Relationships are formatted as in Fig. 1 (influence scores not shown). Habitat variables are open circles, and species are solid squares. New species in 2004 are marked by an arc. Habitats are ordered alphabetically; species are ordered phylogenetically. Node identities, clockwise from first habitat (i.e., 12:00 on the circle), are: altitude, burned heather, cotton grass, common heather, path distribution, grass, vegetation height, non-heather, slope, Canada Goose, Red Grouse, Sparrowhawk, Kestrel, Merlin, Peregrine, Lapwing, Dunlin, Golden Plover, Snipe, Curlew, Redshank, Common Sandpiper, Short-eared Owl, Cuckoo, Skylark, Meadow Pipit, Grey Wagtail, Pied Wagtail, Dipper, Wren, Whinchat, Wheatear, Ring Ouzel, Mistle Thrush, Magpie, Carrion Crow, Twite, Reed Bunting; and, continuing for 2004 networks only, Hen Harrier, Goshawk, Buzzard, Hobby, Little Owl, Stonechat, Raven, Lesser Redpoll.

the others in both years (1990, $W^* = 0.62, 0.61$, for 0.5 and 5 km vs. 0.69, 0.63 for 1 and 2 km, respectively; 2004, $W^* = 0.55, 0.51$ vs. 0.59, 0.59).

A number of species were significantly concordant in importance across years and spatial scales (Fig. 4). We draw attention to Golden Plover, which often had the highest importance; Dunlin, Reed Bunting, Meadow Pipit, Curlew, and Red Grouse were also consistently high. Ranked variable importance indicated that raptors were more strongly connected at larger spatial scales (2 and 5 km) than other bird species in 2004 ($\chi^2 = 9.08$, df = 3, $P = 0.028$ [not tested in 1990 due to the shortage of raptor species]).

Across years variable importance was significantly or nearly significantly congruent (for 0.5, 1, 2, and 5 km, respectively: Kendall's $W = 0.71, 0.60, 0.68, 0.61$; $P = 0.002, 0.08, 0.005, 0.07$). Of particular interest is the change between the two surveys: Little Owl at 0.5 km, Goshawk at 2 km, and Peregrine at 5 km all increased in importance from 1990 to 2004.

Comparison of structures across data sets.—In all cases, the original networks were significantly higher scoring on the original data set than the comparison networks (Appendix C: Fig. C9). The comparison networks recovered from spatial scales more similar to the original data set had higher BSM scores than those from comparison data sets with greater differences of spatial scale (with two exceptions; Appendix C: Fig. C9).

Lasso regression networks.—Compared to the BN results, the lasso regression networks were considerably more dense: even conservatively considering only bidirectional influences, there were 25 links for raw data and 22 for discrete, compared to the BN's 12 (Appendix C: Fig. C10). Where the lasso networks shared links with the BN, however, they tended to agree on sign and

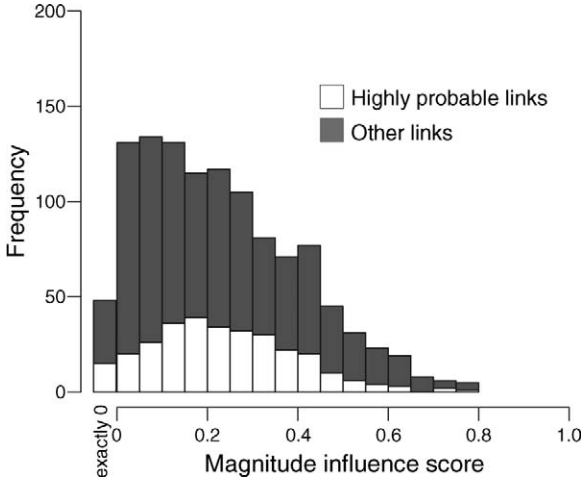
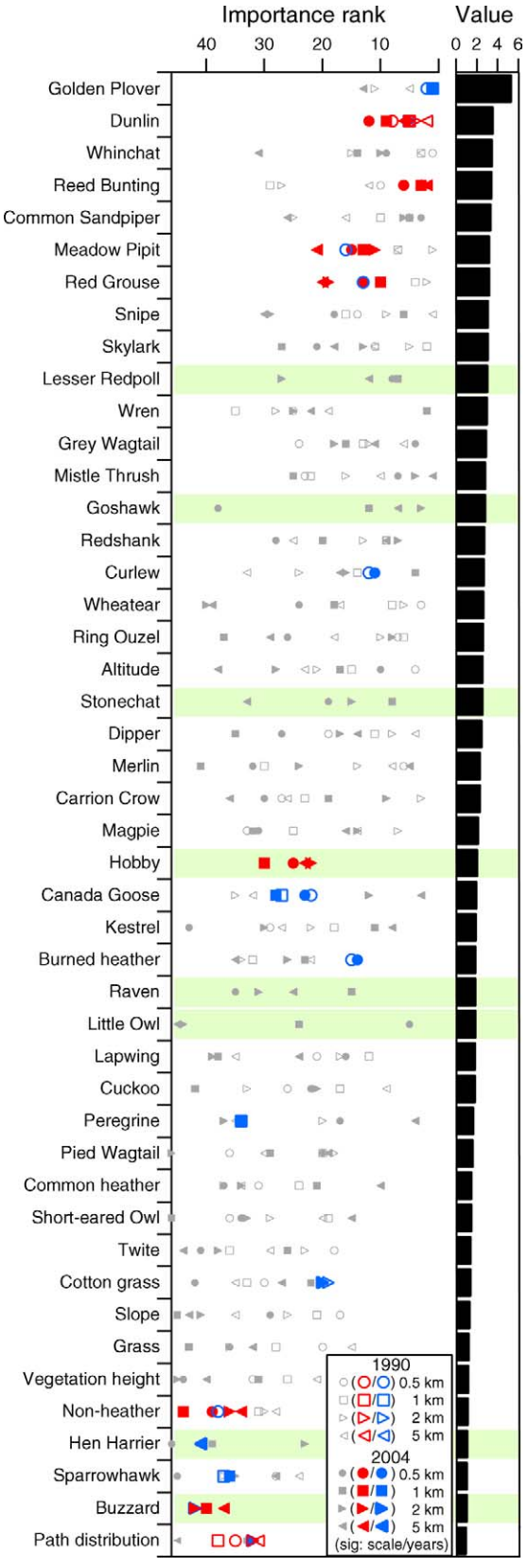


FIG. 3. Magnitude of influence scores. Height of histogram bars shows frequency of influence scores among links; white portion represents highly probable relationships, and gray portion represents other links. Both overall and just considering highly probable relationships, there is a strong skew toward low influence scores. Note that an influence score of exactly 0 does not mean "no influence," and so the bar at this value should be interpreted differently.



relative magnitude of links ($r = 0.85$ raw, 0.94 discrete, $P < 0.0001$); there were only two prominent disagreements: the BNs show a non-monotonic relationship between altitude and path distribution and a negative relationship between vegetation height and common heather, while the lasso networks show negative and positive relationships, respectively.

Discussion

For what we believe is the first time, we were able to use a Bayesian network inference algorithm to recover meaningful networks of functional relationships from ecological data. The networks identified by the algorithm conformed to both the main topological patterns and (with a few notable exceptions) most of the important specific relationships that we expect to find in the upland bird community, giving us confidence in the novel methods and results presented here. Before discussing these results in detail, however, we address the interpretation of these networks. The BN links represent statistical dependence: these are relationships that are predictive in an informative, not causal, manner. The links here reflect samples across a spatial grid; thus, a link indicates significant spatial relationship of species and/or habitat characteristics (e.g., spatial dis- or co-localization of species, positive or negative association in space with values of habitat characteristics). The particular mechanism may vary from pair to pair, ranging from facilitation through trophic interactions, as we see below.

The inter-habitat relationships all matched well-known patterns of co-incidence and exclusion among plant species and environmental variables (Brown et al. 1993). For example, cotton grass is found at high altitude; the recovered networks correctly link altitude to cotton grass with positive influence. Similarly, burned heather is a subset of common heather being managed for Red Grouse. The two are appropriately related with a high positive influence. Additionally, cotton grass grows in wetter areas than common heather, and grass displaces common heather in heavily grazed areas; common heather is linked to both these variables with

←

FIG. 4. The rank of variable importance is plotted for birds and habitat for all spatial scales in both years on the left; bars on the right show average value of variable importance across years and spatial scales: birds and habitats are ordered in descending average importance value. Open shapes represent importance rank in 1990 networks, and solid shapes represent 2004 networks; red designates importance that was significantly concordant across all spatial scales for a given year, and blue indicates importance that was significantly concordant across years. Green background shading denotes birds that were new in 2004. Note that Golden Plover has the highest overall importance and that importance ranks for this species were significantly concordant between years for both the 0.5- and 1-km data sets, suggesting that Golden Plovers are central to the upland bird community network.

PLATE 1. The networks had a recurring triplet of (left) Red Grouse, (middle) Skylark, and (right) Meadow Pipit. Photo credits: John Anderson.

highly negative influence scores. Note that while burned heather is also common heather, and thus shares the same exclusions from grass and cotton grass, the BN only links common heather to the two grasses. This is because burned heather's exclusion from areas of grass and cotton grass is explained entirely through being a subset of common heather. Thus, the BN reflects the direct relationships of common heather with burned heather, grass, and cotton grass, and the indirect relationships between burned heather and the grasses are read from the network structure. Finally, foot paths in the park tend to follow either ridges or valleys, and the BN links altitude to path distribution with an influence of 0.0, correctly indicating the relationship is neither positive nor negative. The fact that these well-known relationships are recovered by the BN is a verification of its ability to find accurate ecological relationships.

In contrast, networks produced from a lasso regression analysis of the same data (Appendix C: Fig. C10) show a nearly fully connected network, including known indirect relationships of burned heather with many variables. This was true even of the lasso analysis on the discretized data, showing that the sparseness of the BN is not due to the loss of detail engendered by discretization, but a consequence of BNs preferentially identifying direct relationships. Thus, the networks produced by BN analysis appear to be more realistically sparse.

Bird–habitat associations are also well known in the uplands (e.g., Pearce-Higgins and Grant 2006), so it is reassuring that we found altitude (known to be one of the strongest influences on the upland bird community) as the most highly connected habitat variable, with burned heather and common heather in second and third place (Fig. 4). Notable, however, is that although slightly over one third of relationships were between birds and habitats, a surprisingly high number of birds had no direct habitat relationships. Most obvious among such absences is relationship between Red Grouse and common or burned heather: a clear indication there are missing habitat associations in these networks. There are several reasons why this may be the case. First, the satellite derived habitat data are subject to error (Pearce-Higgins et al. 2006), reducing the algorithm's performance. Second, the habitat variables are not the complete suite known to influence upland bird distribution. Third, even accurately mapped habitat variables may be a poorer proxy for the cues birds use to determine their distribution than the occurrence of other species that utilize the same resources. Consequently, the network identifies more bird–bird relationships than bird–habitat ones. In particular, in all but the 5-km spatial scale, there is a consistently recurring triplet of Red Grouse, Skylark, and Meadow Pipit, with two or three connections among them (see Plate 1); these two other species may be a more reliable predictor of Red Grouse than heather. We have less a priori knowledge of the likely relationships among bird species, though it is reassuring that raptors are repeatedly linked to known prey species.

We examine global topological features of our networks. First, we find many weak and few strong relationships (Fig. 3). While the precise nature of these relationships are not known, we expect them to be reflective of underlying interactions within the community, thus suggesting similarity to the many weak and few strong interactions expected of stable systems (May 2001, Csermely 2004). Second, we find most inter-bird relationships are positive, as are all relationships between raptors and other birds. Such general positive relationships may reflect direct facilitation (e.g., Dunlin and Golden Plover may benefit through shared nest defense; cf. Larsen and Grundetjern 1997, Haemig 2001, Quinn et al. 2003), use of other species (including competitors and predators) as cues of habitat quality (Stamps and Krishnan 2005), or mediation by unmeasured habitat variables such as food availability. The pattern for raptors matches the most common expectation of predator–prey distribution within single popula-

tions (Sih 2005): predators are found in the presence of their prey. It would be interesting to see if an analysis at a larger scale, whose samples represented population sizes rather spatial distribution, would reveal negative relationships, where absence of predators allowed growth of prey populations.

We also find the emergence of "characteristic scales" of functional relationship. In the bird and habitat networks, significantly more relationships appear at a single spatial scale than expected. Both the comparison of structures across data sets and the fact that variable importance ranks are least similar at the most extreme spatial scales suggests that network structure is more similar at similar spatial scales. All this is expected if relationships operate most strongly at one spatial scale, becoming harder to identify as the scale moves away from the optimum. The characteristic scale associated with raptors was larger than the average characteristic scale of other bird species, as would be expected from wide-ranging top predators. The possibility that relationships may be scale dependent has previously been suggested, but never explicitly identified (cf. Favreau et al. 2006).

Also interesting is the network change between the two surveys. It is clear the ecosystem changed in the fourteen years between surveys: eight new species appeared in the 2004 survey. Although some of these apparent colonizations may represent under reporting (particularly of raptors) in the earlier survey, it also includes some well known colonizations, such as the recolonization of the Raven (Cross 2002). The eight new species had a range of importance, ranging from very low (e.g., Hen Harrier and Buzzard) to medium-high (e.g., Little Owl, Goshawk, Lesser Redpoll), suggesting that species differ in the extent to which they impact and interact with their new community. Additionally, network structure changed significantly between the surveys: networks from one year were always significantly worse at explaining data from the other year. This difference was similar in magnitude to the difference between the one or two closest spatial scales (Appendix C: Fig. C9). Thus, the difference between years indicates a similarly slight shift in the pattern of relationships. Several relationships were found in both years; these were present more often than expected at multiple spatial scales, suggesting relationships with broad characteristic spatial scales were less influenced by network reshuffling between years.

In sum, BN algorithms revealed both known patterns of functional relationship and provided novel insights into spatial and temporal structuring of ecological networks. By producing sparse networks of direct relationships, BNs are likely to be most useful for researchers interested in revealing a minimum set of functional relationships that impacts species distribution. Our application of BNs to ecology takes advantage of spatial effects of relationships, i.e., species distributions are affected by ecological interactions. However, spatial data includes more information than we used, e.g., species which occur in adjacent but not identical locations and other such relative spatial patterns. A method that could incorporate such information may reveal further features of ecosystems, particularly concerning the spatial scale of relationships. Developing a BN algorithm that makes more appropriate use of spatial data sets must be considered a priority to further expand their usefulness in ecology.

Acknowledgments

We are grateful to the Moors for the Future Partnership, Natural England, and RSPB Scotland for allowing access to the data analyzed here. M. Brewer and D. Husmeier discussed the analysis and provided valuable comments on an early draft: we are grateful to them.

Literature Cited

Brown, A., H. J. B. Birks, and D. B. A. Thompson. 1993. A new biogeographical classification of the Scottish uplands. II: Vegetation–environment relationships. Journal of Ecology 81:231–251.

Carr, G., and P. Middleton. 2004. Breeding bird survey of the peak district moorlands. Moors for the Future Report 1. Moors for the Future Partnership, Edale, UK.

Cohen, J. E., K. Schoenly, K. L. Heong, H. Justo, G. Arida, A. T. Barrion, and J. A. Litsinger. 1994. A food-web approach to evaluating the effect of insecticide spraying on insect pest population-dynamics in a Philippine irrigated rice ecosystem. Journal of Applied Ecology 31:747–763.

Cross, T. 2002. Common Raven (Raven) *Corvus corax*. Pages 626–628 *in* C. V. Wernham, M. P. Toms, J. H. Marchant, J. A. Clark, G. M. Siriwardena, and S. R. Baillie, editors. The migration atlas: movements of the birds of Britain and Ireland. Poyser, London, UK.

Csermely, P. 2004. Strong links are important, but weak links stabilize them. Trends in Biochemical Sciences 29:331–334.

Davis, A. J., L. S. Jenkinson, J. H. Lawton, B. Shorrocks, and S. Wood. 1998. Making mistakes when predicting shifts in species range in response to global warming. Nature 391:783–786.

Dunne, J. A., R. J. Williams, and N. D. Martinez. 2002. Network structure and biodiversity loss in food webs: robustness increases with connectance. Ecology Letters 5: 558–567.

Erwin, T. J. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. Coleopterists Bulletin 36:74–75.

Favreau, J. M., C. A. Drew, G. R. Hess, M. J. Rubino, F. H. Koch, and K. A. Eschelbach. 2006. Recommendations for assessing the effectiveness of surrogate species approaches. Biodiversity and Conservation 15:3949–3969.

Friedman, N. 2004. Inferring cellular networks using probabilistic graphical models. Science 303:799–805.

Guisan, A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8:993–1009.

Haemig, P. D. 2001. Symbiotic nesting of birds with formidable animals: a review with applications to biodiversity conservation. Biodiversity and Conservation 10:527–540.

Heckerman, D., D. Geiger, and D. M. Chickering. 1995. Learning Bayesian networks: the combination of knowledge and statistical data. Machine Learning 20:197–243.

Henneman, M. L., and J. Memmott. 2001. Infiltration of a Hawaiian community by introduced biological control agents. Science 293:1314–1316.

Hutchinson, G. E. 1959. Homage to Santa Rosalia or why are there so many kinds of animals? American Naturalist 93:145–159.

Knight, C. G., and C. M. Beale. 2005. Pale Rock Sparrow *Carpospiza brachydactyla* in the Mount Lebanon range: modelling breeding habitat. Ibis 147:324–333.

Krivtsov, V. 2004. Investigations of indirect relationships in ecology and environmental sciences: a review and the implications for comparative theoretical ecosystem analysis. Ecological Modelling 174:37–54.

Larsen, T., and S. Grundetjern. 1997. Optimal choice of neighbour: predator protection among tundra birds. Journal of Avian Biology 28:303–308.

Legendre, P. 2008. Species associations: the Kendall coefficient of concordance revisited. Journal of Agricultural, Biological and Environmental Statistics 10:226–245.

May, R. M. 2001. Stability and complexity in model ecosystems. Princeton University Press, Princeton, New Jersey, USA.

Memmott, J., S. V. Fowler, Q. Paynter, A. W. Sheppard, and P. Syrett. 2000*a*. The invertebrate fauna on broom, *Cytisus scoparius*, in two native and two exotic habitats. Acta Oecologica-International Journal of Ecology 21:213–222.

Memmott, J., N. D. Martinez, and J. E. Cohen. 2000*b*. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. Journal of Animal Ecology 69:1–15.

Montoya, J. M., and R. V. Sole. 2002. Small world patterns in food webs. Journal of Theoretical Biology 214:405–412.

Moya-Larano, J., and D. H. Wise. 2007. Direct and indirect effects of ants on a forest-floor food web. Ecology 88:1454–1465.

Osborne, P. E., G. M. Foody, and S. Suarez-Seoane. 2007. Non-stationarity and local approaches to modelling the distributions of wildlife. Diversity and Distributions 13:313–323.

Pearce-Higgins, J. W., C. M. Beale, J. Wilson, and A. Bonn. 2006. Analysis of moorland breeding bird distribution and change in the Peak District. Moors for the Future Report 11. Moors for the Future Partnership, Edale, UK.

Pearce-Higgins, J. W., and M. C. Grant. 2006. Relationships between bird abundance and the composition and structure of moorland vegetation. Bird Study 53:112–125.

Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecology and Biogeography 12:361–371.

Proulx, S. R., D. E. L. Promislow, and P. C. Phillips. 2005. Network thinking in ecology and evolution. Trends in Ecology and Evolution 20:345–353.

Quinn, J. L., J. Prop, Y. Kokorev, and J. M. Black. 2003. Predator protection or similar habitat selection in red-breasted goose nesting associations: extremes along a continuum. Animal Behaviour 65:297–307.

Sih, A. 2005. Predator–prey space use as an emergent outcome of a behavioral response race. Pages 240–254 *in* P. Barbosa and I. Castellanos, editors, The ecology of predator–prey interactions. Oxford University Press, Oxford, UK.

Smith, V. A., J. Yu, T. V. Smulders, A. J. Hartemink, and E. D. Jarvis. 2006. Computational inference of neural information flow networks. PLoS Computational Biology 2:e161.

Stamps, J., and V. V. Krishnan. 2005. Nonintuitive cue use in habitat selection. Ecology 86:2860–2867.

Stillman, R. A., and A. F. Brown. 1994. Population sizes and habitat associations of upland breeding birds in the South Pennines, England. Biological Conservation 69:307–314.

Tibshirani, R. 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B 58: 267–288.

Yu, J., V. A. Smith, P. P. Wang, A. J. Hartemink, and E. D. Jarvis. 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. Bioinformatics 20:3594–3603.

APPENDIX A

Supplementary methods for data processing, modification of Bayesian network methods for ecological data, and analysis of networks (*Ecological Archives* E091-127-A1).

APPENDIX B

Bayesian network overview (*Ecological Archives* E091-127-A2).

APPENDIX C

Supplementary figures showing detail networks for species and habitat, comparison of scores across data sets, and comparison of Bayesian networks with lasso regression analysis (*Ecological Archives* E091-127-A3).

SUPPLEMENT

R source code (*Ecological Archives* E091-127-S1).