

Bayesian Networks as a novel tool to enhance interpretability and predictive power of ecological models

Edwin Hui^a, Richard Stafford^b, Iain M. Matthews^a, V. Anne Smith^{a*}

^a Centre for Biological Diversity, School of Biology, University of St Andrews, St Andrews, Fife, KY16 9TH, UK

^b Department of Life and Environmental Sciences, Faculty of Science and Technology, Bournemouth University, Poole, BH12 5BB, UK

* corresponding author: anne.smith@st-andrews.ac.uk

Abstract

In today's world, it is becoming increasingly important to have the tools to understand, and ultimately to predict, the response of ecosystems to disturbance. However, understanding such dynamics is not simple. Ecosystems are a complex network of species interactions, and therefore any change to a population of one species will have some degree of community level effect. In recent years, the use of Bayesian networks (BNs) has seen successful applications in molecular biology and ecology, where they were able to recover plausible links in the respective systems they were applied to. The recovered network also comes with a quantifiable metric of interaction strength between variables. While the latter is an invaluable piece of information in ecology, an unexplored application of BNs would be using them as a novel variable selection tool in the training of predictive models. To this end, we evaluate the potential usefulness of BNs in two aspects: (1) we apply BN inference on species abundance data from a rocky shore ecosystem, a system with well documented links, to test the ecological validity of the revealed network; and (2) we evaluate BNs as a novel variable selection method to guide the training of an artificial neural network (ANN). Here, we demonstrate that not only was this approach able to recover meaningful species interactions networks from ecological data, but it also served as a meaningful tool to inform the training of predictive models, where there was an improvement in predictive performance in models with BN variable selection. Combining these results, we demonstrate the potential of this novel application of BNs in enhancing the interpretability and predictive power of ecological models; this has general applicability beyond the studied system, to ecosystems where existing relationships between species and other functional components are unknown.

Keywords

Bayesian Networks, Artificial Neural Networks, Rocky Shores, Variable Selection, Predictive Ecological Model

1. Introduction

Ecosystems are facing multiple pressures, including climate change, biodiversity loss, habitat loss and pollution, on global, regional, and local scales (Steffen et al., 2018). Such pressures have significant impact on population dynamics, community structure, and ecosystem function. In many ecosystems if pressures exceed a threshold, or tipping point, the changes may lead to an ecological regime shift and potentially a new alternative stable state (Petraitis et al., 2009; Petraitis and Dudgeon, 1999). Regime shifts are large and sudden changes where an ecosystem undergoes a step change to another state that can last for substantial periods of time (typically decades for most ecosystems), and potentially indefinitely (Folke et al., 2004; Scheffer et al., 2001). This is detrimental to ecosystems for two reasons. Firstly, post-regime states may be highly stable themselves. Such shifts entail changes in internal dynamics and feedbacks of an ecosystem that often make it impossible to reverse to its original state, even if efforts are made in removing the drivers that lead to the shift (Scheffer et al., 1993). Secondly, alternative regimes are normally 'poorer' in both biodiversity and provision of ecosystem service (Hawkins et al., 2015).

Given this, it is of paramount importance to be able to understand, and ultimately to predict, the response of ecosystems to disturbance, including resilience of the ecosystem and potential new stable states. However, understanding such dynamics is not simple. Ecosystems are a complex network of interactions: species exists as part of an ecological community that is dictated by interactions with prey, predators and competitors. Therefore, any change to a population of one species will have some degree of community level effect (Henneman and Memmott, 2001; Stafford et al., 2013).

Computational inference of complex networks presents an efficient route to reveal complex interactions such as those within an ecosystem, and have been demonstrated to work on some natural complex systems. For example, the recent applications of Bayesian Networks (BNs) have shown success in recovering gene regulatory networks from gene microarray data (Chen and Mar, 2018; Friedman et al., 2000; Hecker et al., 2009). BNs have also seen successful applications in ecology, where recovered networks from observational data corresponded well to known species and habitat interactions (Milns et al., 2010; Mitchell et al., 2021; Trifonova et al., 2015). Given that it has been established that understanding these interactions is crucial in understanding the impact of large scale disturbances such as climate change on ecological systems (Pearson and Dawson, 2003; Proulx et al., 2005), BNs therefore offer a novel way to reveal ecological network structures within stable states using species counts that are relatively easy to obtain.

Another key strength of BNs is the ability to at least semi-quantify the strength of interactions between variables (or species). Using this information, it follows that relative variable importance with respect to a certain variable of interest can be inferred from the revealed network. Knowing the relative strength of certain species in relation to others is invaluable in the field of ecology. For example, knowing the strength of interactions between predator, prey and competitors would give a clear understanding of how ecological communities are structured and regulated.

The combination of these two features, network structure and interaction strength, has the potential to serve as a novel variable selection tool in the field of machine learning. The revealed structure can identify relevant variables in relation to a target variable. This could guide the training of various predictive models such as multiple linear regression (MLR) models, one of the most popular models in ecology. However, a key limitation of using MLR models or derivatives of them such as generalised linear models (GLM) as a predictive tool for ecosystem dynamics is their limited ability to deal with non-linear relationships between the dependent and independent variables (Brosse et al., 1999; Gevrey et al., 2003; Laë et al., 1999). A potential alternative to these approaches is the use of artificial neural networks (ANNs). ANNs have gained increasing attention in ecological modelling in the last decade, with the main reason being their ability to detect patterns through complex, non-linear relationships (Jeong et al., 2001; Kroodsmas et al., 2018; Mac Aodha et al., 2018; Pereira et al., 2019). Therefore, this makes ANNs an attractive alternative to conventional regression models. However, ANNs lack easily-interpretable coefficients as found in MLR models. Thus, use of a BN variable selection tool would complement an ANN predictive model by providing interaction strengths for the selected variables.

In this paper, we aim to evaluate the potential usefulness of BNs in two aspects. Firstly, we apply BN inference on species abundance data from a rocky shore ecosystem in Scotland. Rocky shores are systems with well documented relationships between species and this allows us to test the ecological validity of the revealed network against experimentally derived field data. Secondly, we evaluate BNs as a novel variable selection method to train an ANN for each component species of the network. To evaluate the effectiveness of BN as a variable selection method, we compare the performance of the ANN with and without the BN-based variable selection. Finally, to benchmark our approach against previous methods, we compare the performance of the ANN with variable selection against a generalised linear model (GLM, a type of MLR model) with variable selection, where variable selection has been performed by the BN.

2. Methods

2.1 Field Methods

Our study site was two continuous sections of rocky shore of 10m in length at East Sands, St Andrews, Scotland (56°20'04.0"N 2°46'23.2"W). Both sites were at a tidal height of 2.9 m above chart datum. These sites were selected based on initial inspection of community structure to ensure that both macroalgae and barnacle stands were present, but were otherwise haphazardly selected from other potential sites at this height on the shore.

All sampling occurred in May 2020. Sampling only occurred during low tide time periods. Fifty haphazardly placed 50x50cm double strung quadrats were placed at each site. Grazer count (for *Littorina littorea* (littorinids) and *Patella vulgata* (limpets)) were obtained for each quadrat. Percentage cover estimates for barnacles (*Semibalanus balanoides* and *Chthamalus stellatus*), macroalgae (*Ascophyllum nodosum* and *Fucus vesiculosus*) and microalgae (Biofilm) were obtained by photographing the quadrat frames using an iPhone XR. Estimates were then made of percentage cover for each species present. Other grazer species were extremely rare, and accounted for < 2% of grazers found. This was considered unlikely to affect our results, so are not included in subsequent analysis. This sampling covered approximately 50% of the area being considered, which has been demonstrated to be sufficient to capture the variation and details of even the most patchily distributed organisms on the rocky shores (Stafford, 2002; Stafford et al., 2015). The research data underpinning this publication can be accessed at <https://doi.org/10.17630/f2b69f88-efb7-43a1-96e9-70012256a752> (Hui et al., 2021).

2.2 Data pre-processing

Data analysis was carried out using R 4.0 (R Core team, 2020) to prepare the data for further analysis. It is traditional knowledge in the field of machine learning that the more data that any machine learning algorithm has access to, the more effective it can be (Perez and Wang, 2017). The problem with small datasets is that models that are trained with them do not generalize well beyond the original dataset. Hence, these models suffer from overfitting.

To combat against this problem, and with the knowledge that our sampling sufficiently captured the site variation (Stafford, 2002; Stafford et al., 2015), we applied a uniform data augmentation method. This process simply takes each data point, in this case a single cell of percentage cover (0-100%) or grazer count (0-42 individuals), and returns an additional point that either adds or subtracts a value from a uniform distribution of ± 3 . This would take on the form as

$x + \text{runif}(1, -3, +3)$, with values <0 reset to 0. This process was repeated until 400 data points was reached. When partitioning the data into training and validation sets (in the neural networks), this process was applied only to the training data partition.

3. Modelling

3.1 Bayesian networks

Bayesian networks consists of a graphical representation of the joint probability distribution among a set of variables $X = \{X_1, \dots, X_n\}$. The graphical representation, in the form of a directed acyclic graph (DAG), represents the dependencies and conditional relationships among a set of variables. This can be defined with the pair $\langle G, \theta \rangle$. G represents a directed graph with nodes that correspond to the variables X_1, \dots, X_n , and links between nodes indicate statistical dependency, where the child variable is dependent on its parents. The graph G and the joint probability distribution of θ are connected to together by the Markov condition property: variable X_i is conditionally independent from all non-descendants, given its parents in G , $Pa(X_i)$. Lastly, the parameters of θ specifies the probabilistic relationship of each node to its parents, $P(X_i|Pa(X_i))$. Given these qualities, the joint probability distribution that is described by $\langle G, \theta \rangle$ is therefore:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i|Pa(X_i)) \quad (1)$$

3.1.1 Learning Bayesian networks

Banjo 2.2 (<https://users.cs.duke.edu/~amink/software/banjo/>; Yu et al., 2004, Smith et al., 2006) was used to reveal the Bayesian networks. Due to the discrete nature of the learning algorithm, the dataset had to be discretized. To this end, all variables were discretized into 3 bins using quantile discretization. The variables were mapped into 3 bins as follows:

Limpet: [0, 9] → low, [10, 29] → medium, and [30/42] → high

Littorinid: [0, 4] → low, [5, 8] → medium, and [9/21] → high

Barnacle: [0, 6] → low, [7, 49] → medium, and [50/88] → high

Macroalgae: [0, 5] → low, [5, 8] → medium, and [9/21] → high

Microalgae: [0, 10] → low, [12, 50] → medium, and [53/80] → high

The number of bins were selected due to the following reasons: 1) it is intuitive as it reflects 'low', 'medium', and 'high' ordinal states; 2) to minimize the quantity of conditional probabilities while simultaneously allowing for non-monotonic relationships to be revealed, and 3) discretization into 3 bins has been shown to be optimal in simulation studies and had success in previous applications of Bayesian networks in ecological systems (Yu et al., 2004; Milns et al., 2010). To search for the best networks, Banjo uses a heuristic search to identify top scoring networks based on the Bayesian Dirichlet equivalent scoring metric. Banjo was run using settings: greedy search, using random local moves, with an equivalent sample size of 1. Upon completion of a search, Banjo returns a single top scoring network along with influence score for each link, which ranges from -1 to 1, representing the direction and magnitude of influence, with a score of exactly 0.0 representing non-monotonic (e.g. hump- or U-shaped) influence.

3.1.2 Using Markov blanket from Bayesian networks for variable selection

A Markov blanket (MB) of any target variable T , $MB(T)$, is the minimal set for which the conditional independence relationship I between X and T holds such that $I(X; T | MB(T))$, for all $X \in V - \{T\} - MB(T)$ in a variable set V . Therefore, the MB of a given target variable includes its parent nodes, child nodes, and other parent nodes of the children. This condition renders variable T statistically independent from all the remaining variables of a BN, given the values of the variables in the Markov blanket. Putting all the above together, we can describe the relationship between $\langle G, \theta \rangle$ as follows: a BN graphical structure of G is faithful to a joint probability distribution θ over a variable set V if and only if every dependence entailed by the graph G is also present in θ . This formal definition is the basis of our proposed variable selection method: we identify the Markov blanket of specific variables within the revealed Bayesian network, and these variables will be selected for training the predictive models.

3.1.4 Learning variable importance from Bayesian networks

To quantify variable importance from each component variables of the revealed Markov blankets, we utilize Banjo's influence score and mutual information. Upon revealing the Markov blanket for each component variable within the Bayesian network, we parse each respective Markov blanket through Banjo to obtain the influence scores for each variable with regards to their target variable.

To complement the influence score, which shows the direction and magnitude of relationship, we also compute the mutual information using the R package *entropy v1.3.0* (Hausser and Strimmer,

2009). The mutual information of two random variables is a measure of mutual dependence between two variables, where it quantifies the ‘amount of information’ obtained about one random variable through the observation of another random variable. For example, the mutual information (I) between variables X and Y can be defined by:

$$I(X, Y) = H(X) - H(X|Y) \quad (2)$$

where H represents the entropy of a given variable.

This is then equivalent to the following:

$$I(X, Y) = \sum_{y \in Y} p(y) \sum_{x \in X} p(x|y) \log_2 \frac{p(x|y)}{p(x)} \quad (3)$$

This allows mutual information between a target variable and any possible predictor to be computed. As a direct result, we can then infer which variable provides the maximum information gain, and thus allows us to gauge variable importance in relation to a target variable, in a manner complementary to the influence score.

3.2 Artificial Neural Networks

An Artificial Neural Network models the relationship a set of input and output signals using a model that is derived from animal neural system mechanisms (van Wijk and Bouten, 1999).

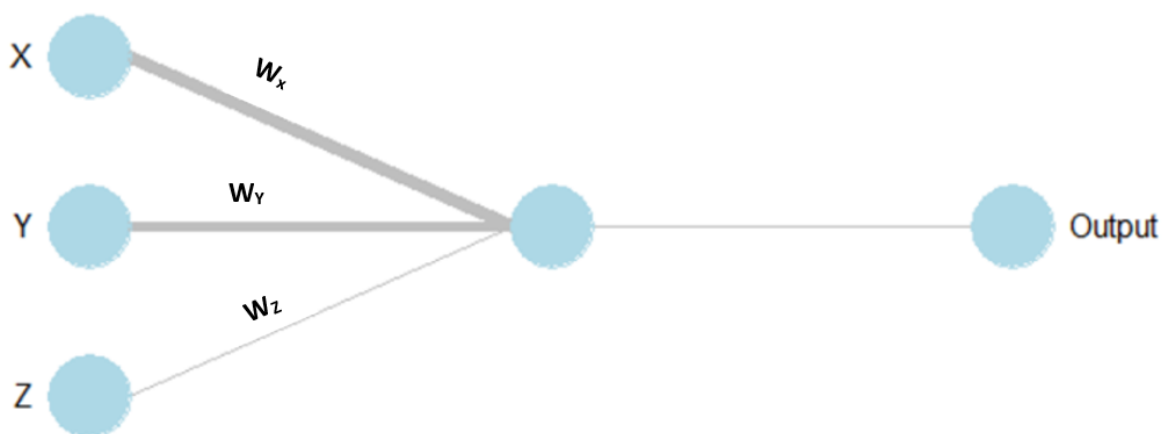


Figure 1: A single artificial neuron. Inputs X , Y , and Z are weighted by connection weights w_x , w_y and w_z , respectively, and combined within the neuron via its activation function $f(x)$ to result in the Output.

The model of a single artificial neuron is comparable to the biological model. Biological neurons receive incoming signals via their dendrites. This impulse is then weighted according to its relative importance, and as the cell body begins to accumulate weighted signals, a threshold is reached and the cell fires off an output signal. This process is essentially mirrored in artificial neurons (Fig. 1), where input signals are received by the dendrites (variables X, Y, Z) and are combined into the output signal (Output). Each signal is weighted (w_x , w_y and w_z) according to its importance and then summed up by the cell body, and the signal is passed according to an activation function. Therefore, the output an artificial neuron with n inputs can be expressed as the following:

$$y(x) = f(\sum_{i=1}^n w_i x_i) + b \quad (4)$$

where w_i represents the connection weights which allows each of n inputs (x_i) to contribute to the sum of input signals. The net sum is used by the activation function $f(x)$ and the resulting signal $y(x)$, including the addition of a bias term (b), is the output. Here, we use the logistic activation function defined by: $f(x) = \frac{1}{1+e^{-z}}$. The main advantage of using this is that it provides a smooth gradient of outputs, preventing ‘jumps’ in output values.

3.2.1 General network architecture

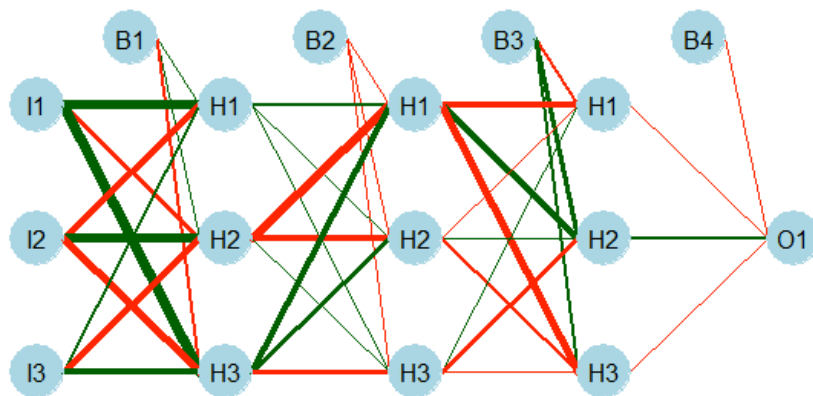


Figure 2: Generic structure of a feed-forward artificial neural network. The first layer represents the input neurons (I1-I3). The next 3 layers represent the hidden layers, and their associated hidden neurons. The final layer represents the output neuron (ON1). Bias terms per layer are represented by B1-4. Line represents connections between neurons, where the weights are represented by the thickness of the lines and the colors represent the effect, where green represents a positive effect and red representing negative.

Typically, an ANN is composed in a manner where every neuron of one layer connects with all neurons in the next layer. The first layer is the input layer, which in this case represents input from each independent variable (Fig. 2). This layer receives data and transmits data to the next layer, which is the hidden layer. Each cell in the hidden layer acts as a processing neuron which receives information from different input neurons. These cells sum up the input signals and processes the data according to an activation function. This in turn produces a signal for the next layer. The next layer can either be another hidden layer, or an output layer, which sums up all the incoming signals to produce a response, which in this context will be the value of the dependent variable. The connection between neurons vary by magnitude and direction: they can either be positive or negative, and they vary in high or lower effects. This is known as the 'connection weights', where the effect of one neuron on the next may be positive or negative depending on the sign of the weight (Olden and Jackson, 2002). Additionally, each hidden layer and output layer has a bias term added, which are numeric constants that allow the value at indicated neurons to be shifted upwards or downwards (Lantz, 2013).

3.2.2 Training algorithm: Gradient based optimisation

Recall that each neural network output transforms input data as defined by equation 4. In this expression, w and b are the weights and biases which are the trainable parameters, and these weights contain the information learnt by the network from exposure to training data. Initially training starts with a random initialization of w and b , which allows us to obtain predictions of $y(x)$. The algorithm then computes the loss, which is the degree of mismatch between the predicted outputs and the actual outputs. Given the degree of loss, the algorithm then gradually adjusts these weights in a way that slightly reduces the loss. This becomes a training loop, which is repeated until a very low loss on training data is achieved.

Here, the activation function is crucial. Because the sigmoid activation function is differentiable, the training algorithm uses the derivative of each neurons activation function to identify the gradient in the direction of each training weight. Given the gradient, it allows the algorithm to measure how much the loss can be reduced for a change in weight. This process is known as the stochastic gradient descent and is the backbone of the training process in ANNs. In this paper, we utilize the backpropagation algorithm to train our ANNs.

3.2.3 Training Artificial Neural Networks

R was used to train the ANNs. All data points had to be modified as neural networks work best when the input data are scaled to a narrow range near 0 (Lantz, 2013). We defined our own min-max normalization function:

$$g(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (5)$$

This function was then applied to all variables independently to scale our data.

To train the model, we used the R packages *caret v6.0-88* (Kuhn, 2020) and *neuralnet v1.44.2* (Fritsch et al., 2019). Our dataset was randomly partitioned into 70% training and 30% validation. We used *caret's* maximum of 3 hidden layers, of which each layer is allowed to have 1-4 neurons. Therefore, to tune these hyperparameters, we performed 10 searches for each layer starting from layer 1. During each iteration, *caret* performs backpropagation to determine the most optimal number of neurons per hidden layer using 10-fold cross validation. The root mean square error (RMSE) is calculated and the model with the lowest RMSE is selected. This process is then repeated until the optimal number of neurons for all 3 layers have been learnt. For each component species of the revealed BN, we trained two ANNs: (1) where every other variable acts as the predictor (no variable selection); and (2) where only the variables within the Markov blanket of the target variable are used as the predictor (variable selection). To test our model performance, we utilized the 30% validation set as input for each model, and then measured the Pearson correlation between the model predicted output and the true value. We replicate this process 10 times for both no variable selection and for variable selection.

3.2.4 Learning Variable Importance from ANN

To quantify variable importance from our ANN models, we utilize Olden's connection weights algorithm (Olden et al., 2004). This method calculates importance as the summed product of the raw input-hidden and hidden-output connection weights between each input and output neuron (Beck, 2018). To apply this algorithm, we use the *NeuralNetTools v1.5.2* package (Beck, 2018).

3.2.5 Training GLM models

R was used to train the GLM models. The same training and validation data used for training ANNs were used to train the GLM models. The data did not require min-max normalization. To train the

GLM models, we used the glm function, with the Poisson link function, along with the *caret* package. For each model, we utilized a 5-fold cross validation training procedure. To test model performance, we use the same procedure outlined in the previous section.

4. Results

4.1 Bayesian Networks

The network shown in Fig. 3 represents the revealed BN of rocky shore species. There was a strong negative relationship between macroalgae and barnacles. Additionally, positive links were found between grazers (limpets and littorinids) and macroalgae and microalgae. A non-monotonic link (0 influence score) was found between limpets and barnacle.

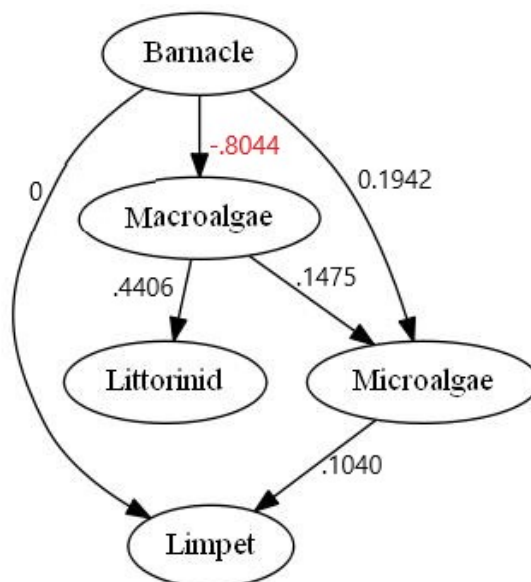


Figure 3: Bayesian network showing the revealed species interaction network of a typical rocky shore community. Values represent influence scores.

4.2 Markov blankets and variable importance

Fig. 4 represents the Markov blanket of each component species of the Bayesian network shown in Fig. 3, along with the influence scores and mutual information. The Markov blanket for each species shows which species are the most important in predicting the specified target variable. The influence scores represent the direction and magnitude of relationship between the predictor and target variables, while the mutual information shows the extent to which having knowledge of a predictor variable reduces the uncertainty about the target variable. Both influence score and mutual

information was the strongest when sessile species were used to predict each other, where barnacle was the most important in predicting macroalgae and vice versa.

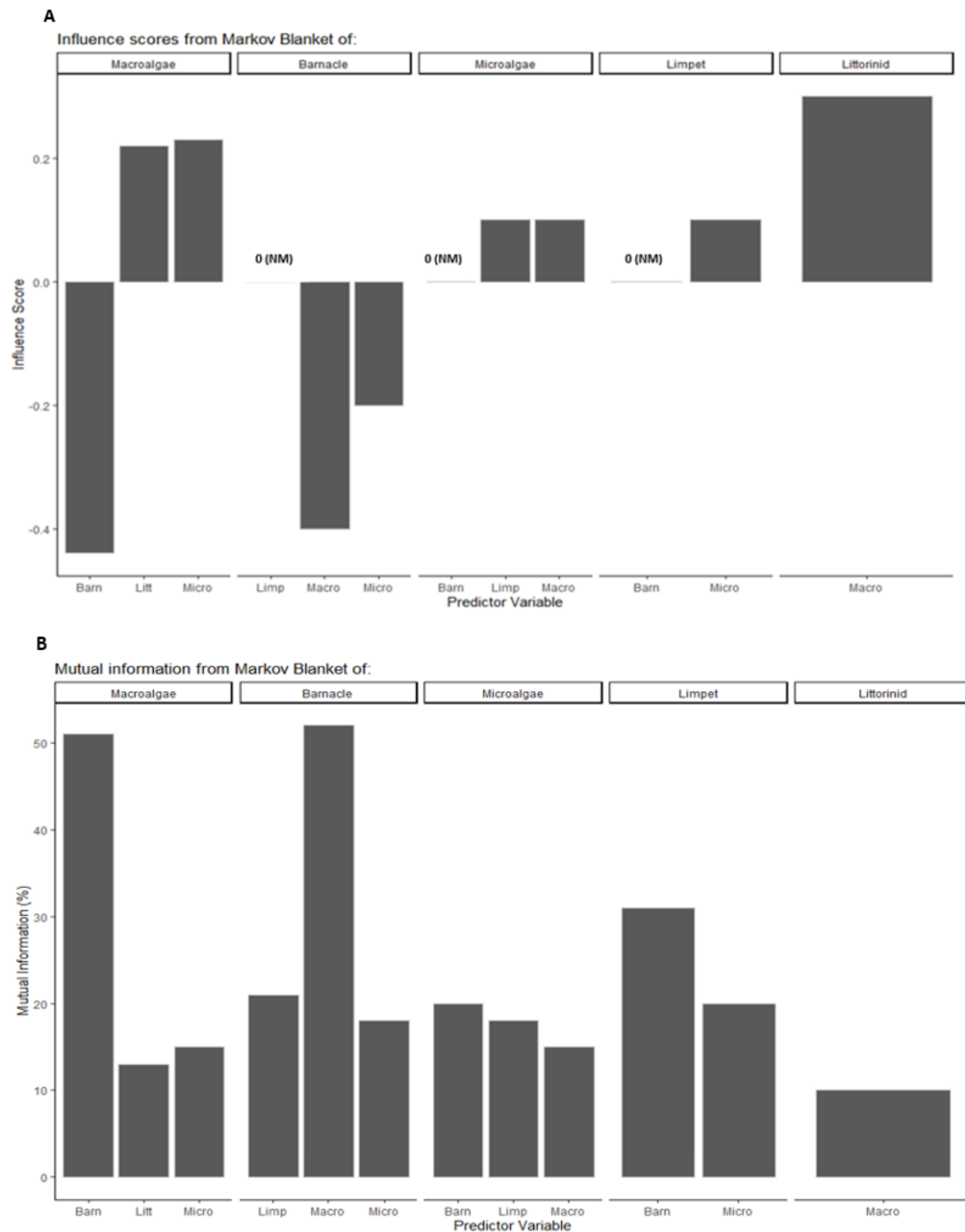


Figure 4: Markov blanket of each target variable (across top of plots) with influence scores (A) and mutual information (B) of each predictor (x-axes). Influence scores show the magnitude and direction of relationship between predictor variables and target variable. Influence scores of '0' represent a

non-monotonic (NM) relationship. The mutual information is the ‘amount of information’ obtained about the target variable through the observation of another predictor variable.

4.3 Artificial Neural Networks

For each component species in the BN (Fig. 3), we trained one ANN using all other variables as a predictor (no variable selection), and another ANN using only the variables within the Markov blanket of the selected variable (variable selection). There were significant improvements in model performance with variable selection in macroalgae and barnacle (Figure 5: with predictor variables as defined in Fig. 4; significance determined by t-tests: macroalgae: $t_{(17.442)} = -16.903$, $p = 2.96 \times 10^{-12}$; barnacle: $t_{(-15.377)} = -15.377$, $p = 1.37 \times 10^{-8}$). On the other hand, no significant difference occurred with variable selection for microalgae, littorinids or limpets (microalgae, limpet and littorinid; $p > 0.05$).

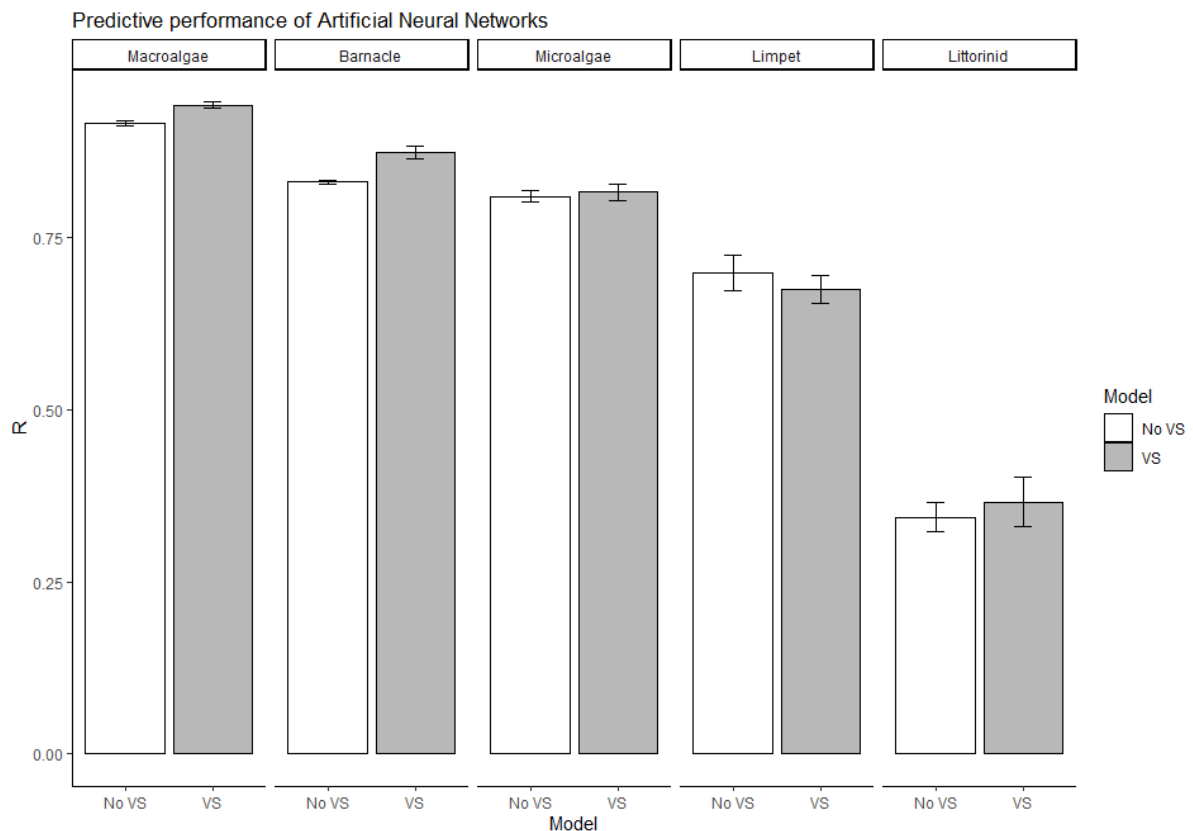


Figure 5: Comparison of model performance between ANN with variable selection and without variable selection, for predicting each variable across top. White bars = ANN with no variable selection (No VS), Grey bars = ANN with variable selection (VS). Error bars represent standard deviation. R represents Pearson correlation coefficient between model predictions and test data.

4.4 Variable Importance from ANNs with variable selection

Olden's connection weight algorithm, when applied to each model, revealed which species were the most important in predicting the specific target variable, where it shows the strength and direction of this relationship. This showed similar patterns with the results from our BN influence scores (see Fig. 4), where the relationships were the strongest when sessile species were used to predict each other. However, the level of uncertainty for specific variables was large between model fits for the same network architecture.

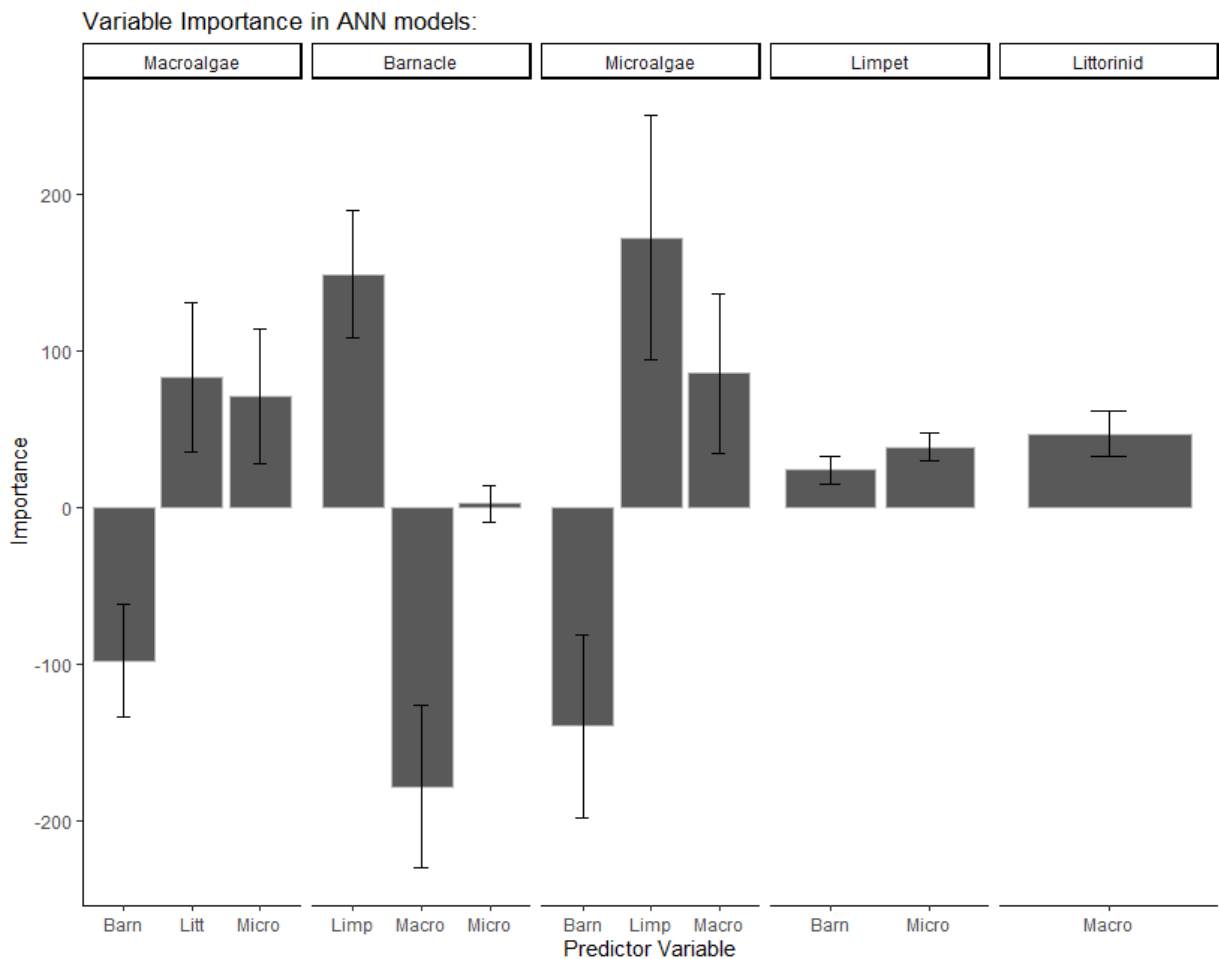


Figure 6: Variable importance derived from ANN models using Olden's connection weight algorithm. Importance for predictor variables (x-axis) shown for each set of models predicting each variable in turn (top). Error bars represent standard deviation.

4.5 Comparing ANN with GLM models

A total of 5 GLM models with variable selection were trained. The predictive performance of these models followed a similar trend observed in the ANNs, where performance was better for sessile species (barnacle, macroalgae and microalgae) compared to the grazers. However, the predictive performance of ANN outperformed GLM models across all species (Fig. 7). There were significant improvements in model performance in ANNs compared to GLMs for macroalgae, barnacle, microalgae and littorinid models (significance determined by t-tests: macroalgae: $t_{(10.012)} = 10.012$, $p = 1.184 \times 10^{-9}$, barnacle: $t_{(10.999)} = 10.827$, $p = 3.326 \times 10^{-7}$, microalgae: $t_{(15.284)} = 6.5204$, $p = 8.833 \times 10^{-6}$, littorinid: $t_{(17.927)} = 4.4089$, $p = 0.00034$). However, there was no significant difference in model performance for in the Limpet ANN and GLM models (Fig. 7: limpet, $p > 0.05$).

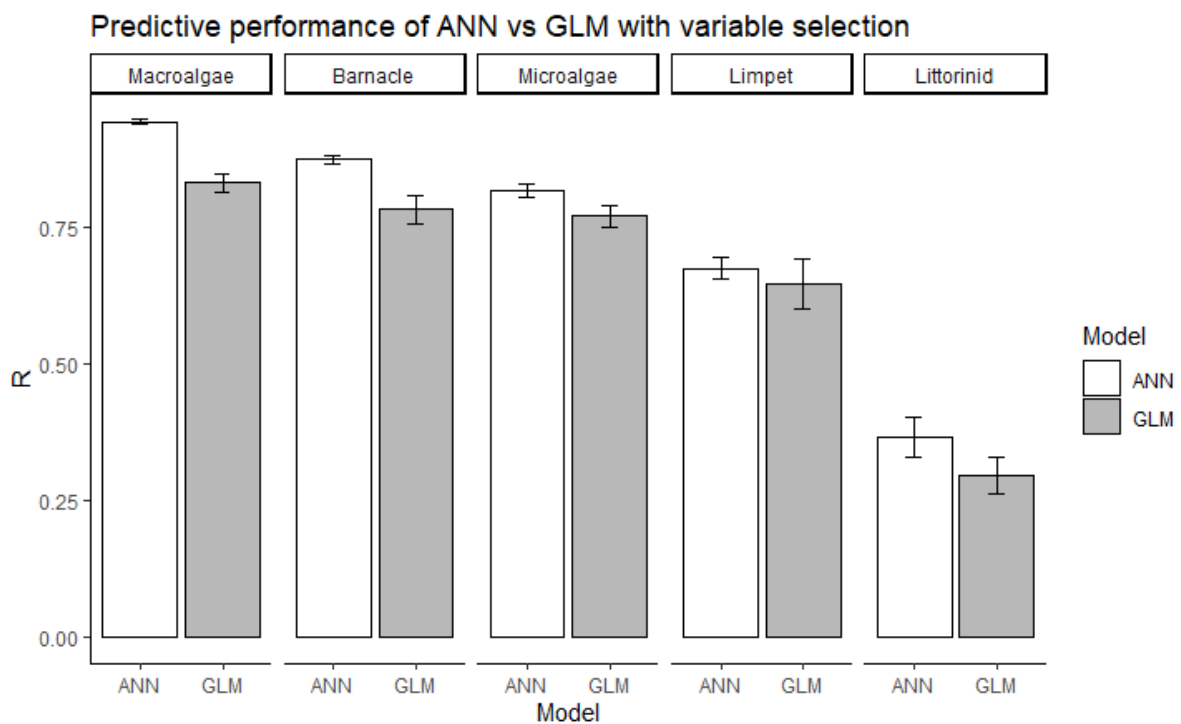


Figure 7: Comparing predictive performance of ANN and GLM models for each target variable (across top). Error bars represent standard deviation. R represents Pearson correlation coefficient between model predictions and test data.

5. Discussion

The application of Bayesian networks to rocky shore ecosystems predicted relationships between species well, and provided relative weights to indicate the importance of the interactions. Although not those expected to be obtained immediately after experimental manipulations, these results match what one would expect from a 'static' rocky shore system (i.e. those which are in a 'stable state' rather than those adapting post experimental disturbance). The use of BNs for variable selection in ANNs demonstrated improved model performance in some cases, and addresses some ecological concerns around the 'black box' nature of ANNs. Finally, the predictive power of ANNs was shown to be greater than GLM for the rocky shore community we examined. These points are examined in more detail below.

5.1 Bayesian networks revealed known functional relationships

The BN revealed some inconsistency with prior experimental knowledge of competitive and grazing relationships. However, the BN revealed important specific relationships that one would expect to find on the rocky shores, given relationships between variables. From a BN conducted outside of experimental manipulations, as was the case in this study, i.e. within a given stable state, these links represent statistical dependencies, where links between variables are predictive in an informative manner, and not causality, which is often obtained through experimental manipulations.

The strong negative relationship found between macroalgae and barnacle reflects the alternative stable states that have been documented in previous literature (Petraitis et al., 2003; Petraitis and Dudgeon, 1999), and the competition for space found on rocky shores (Raffaelli and Hawkins, 1999). Macroalgae stands and barnacle stands represent two different states, where both sessile species compete for space on the rocky shore. Therefore, it follows that having knowledge about the presence of barnacles is informative of the presence of macroalgae, as a high level of barnacle would suggest that space has been exploited, therefore leaving no room for macroalgae to establish (and vice versa). Here, it should be noted physical factors such as wave exposure and shore angles have been demonstrated to influence the distribution of macroalgae and barnacles as well. For example, biomechanical analysis of wave action has shown that dislodgement of macroalgae in land-facing sites were far less likely than sea-facing sites (Jonsson et al., 2006). Therefore, in seaward facing sites that are generally more exposed to wave action, patches with higher proportion of barnacles may not simply just be attributed to a exploitation of space, but also the physical factors that limit macroalgae from occupying these areas.

Grazing relationships appear to be captured in the model as well. Positive links were found between littorinids and macroalgae. This was consistent with prior knowledge as macroalgae has been documented to provide food (via spores) and shelter for littorinid snails, whilst simultaneously acting as a buffer from wave action (Norton et al., 1990). However, it should be noted that the links found reflect a static 'snapshot' of a typical rocky shore. In a dynamic system with the consideration of time, we would also expect littorinids to reduce the level of establishing macroalgae (Hidalgo et al., 2008). Therefore, it could also be argued that there should be a negative link found between littorinids and macroalgae. However, sampling occurred at a time where macroalgae had already established, and therefore acted as a refuge for littorinids. This led to a positive link being recovered: which was consistent with expectations for quadrats with established macroalgae (Norton et al., 1990).

This was the same with the relationships between limpet and microalgae. Generally, limpets graze on microalgae patches, which should lead to a negative link. However, they may be also attracted to areas with microalgae, thus leading to a positive link (Jerkanof, 2006; Nicotri, 1977). The latter appears to be what was revealed by the static Bayesian network, where limpets were generally associated with areas microalgae, rather than from experimental studies, where systems are settling into a new state following an experimental disturbance.

The above relationship should hold between limpets and macroalgae as well – where limpets are expected to be attracted to areas with macroalgae, as they have been documented to graze on macroalgae patches (Arrontes et al., 2004; Davies et al., 2007). At the same time, studies have demonstrated that limpets control growth of established macroalgae, where increased limpet density around macroalgae patches decreases the breaking force of macroalgae, thereby increasing the vulnerability of macroalgae patches to wave induced breakage (Davies et al., 2007). Therefore, following the 'static system' argument, we should expect at least one of the two relationships (negative or positive) between limpets and macroalgae to be revealed. However, there was no link found. Additionally, it is unclear why there was a non-monotonic link between barnacles and limpets. While this is not an issue for the Bayesian network alone as expert knowledge can be used to validate links, this becomes a problem when using this specific network structure to perform variable selection to train a model to predict limpet abundance. This will be discussed in section 5.3.

5.2 Limitations of Bayesian network

It becomes apparent that Bayesian network learning algorithms have revealed many known patterns of the functional relationship between species in a static ecological network. Additionally, it was able to provide a quantifiable measure of the strength of interactions between species community. Both are invaluable in the study of ecology. It should be noted, however, that the system under study was relatively static and within a stable state when we measured species count/% cover. Therefore, the recovered network represents a static snapshot of ecological interactions at a single time-step. Therefore, the next step to further our understanding in ecological networks would be implementing the consideration of time, and the effect of manipulations of grazer or producer density. This is crucial as not only will this allow the effects of competition and predation to be modelled on a dynamic time scale, but it also opens opportunities to study how ecosystems could recover from disturbance and how relationships change between potentially different states. Finally, such a network can be validated alongside known species relationships derived from considerable literature on manipulative experiments. This is especially important in the wider ecological context, where having the tools to predict the onset of a regime shift is becoming increasingly important (Folke et al., 2004; Stafford et al., 2013).

5.3 Artificial Neural Networks show improvements with BN variable selection

There were general improvements to model performance in ANN models with variable selection compared to models with no variable selection. This difference was significant for barnacles and macroalgae. From our models, there was a clear distinction of predictive performance of sessile species (barnacles and macroalgae) and mobile species (limpets and littorinids), where models of sessile species performed better than models of mobile species. From a causal perspective, this would make sense. Given the barnacles and macroalgae confer alternative assemblages on the rocky shore, the presence of one sessile species has strongly affects the presence of another. However, it should be noted that physical abiotic factors such as wave exposure and site angle also have a strong influence on the distribution of these sessile species (see section 5.1). While the absence of these physical factors did not affect the performance of the barnacle and macroalgae models, this becomes problematic when modelling the distribution of grazers. Firstly, grazers are mobile. Therefore, despite a relevant biological relationship between macroalgae and littorinids, there is no way for the model to explain the presence of littorinids in areas with no macroalgae. In these instances, factors such as crevices and pits have a significant role in the distribution of littorinids (Chapman and Underwood, 1994; Seuront and Ng, 2016). For example, it has been demonstrated

that crevices play a key role in the survival of littorinid snails, as they serve as refuge against predators (Catesby and McKillup, 1998). Additionally, crevices play a role regulating thermal stress (Seuront and Ng, 2016). This would explain the poor performance in littorinid models, where despite revealing the correct variable via Markov blanket, the predictive performance was still very poor as it could not account for littorinids being present in quadrats with low % cover of macroalgae, where physical abiotic factors such as crevices may have driven underlying littorinid aggregations. This likely led to overfitting of ANN models for littorinids, especially on randomly selected training and test data.

The lack of physical abiotic factors may have contributed to the drop in performance for limpet models with variable selection as well. Factors such as exposure to wave action are equally important in the distribution of limpets (Thompson, 1980), where it has been demonstrated that wave action has a significant effect on growth rates and mortality rates of limpets. Another potential reason for the drop in the limpet models could be due to the original Markov blanket revealed in the BN, where only microalgae and barnacles were selected. Given that limpets have been documented to reside in both macroalgae and barnacle stands, one would expect to find links between the two variables (Thompson, 1980). Here, the absence of this relationship may potentially be attributed to a site-specific factor. Sampling occurred during late spring. While this usually entails an important growth period for macroalgae, an unexpected storm event or potential out of season factors may have caused a proportion of macroalgae patches to die off. While limpets were largely unaffected, the expected relationship between macroalgae and limpets was lost, as most of the limpets in my samples were closely associated with barnacles and microalgae.

Here, while species count data was sufficient in training an accurate model for sessile species, physical abiotic factors would be required to fully capture the variation in grazers. Factors such as exposure to wave action, desiccation, and shore angle, along with biotic factors such as predation and competition all have a dynamic role in structuring intertidal communities (Raffaelli and Hawkins, 1999). Additionally, it should be noted that these models reflect a static snapshot of an intertidal community – therefore, when trying to model more complex processes with the consideration of time, this issue of missing physical factors would likely lead to even more inaccurate models.

5.4 Variable importance from BNs show more consistent and robust results than ANNs

The variable importance obtained from the BN and ANN were relatively consistent, where both models showed similar patterns of sessile species being the strongest predictor when used to predict

each other. There were instances where the BN influence scores would show a non-monotonic link, whereas the ANN connection weight algorithm would show a strong effect. Here, we argue that the information from the BN is more reliable than ANN. Firstly, the BN comes with a quantifiable metric of interaction strengths via the influence score, which shows the direction and magnitude of relationship. This can then be complemented with mutual information which quantifies the 'amount of information' obtained about one variable through the observation of another variable. This becomes particularly useful when the influence scores show a non-monotonic relationship, as mutual information can provide further information about the nature of the relationship between two variables.

On the other hand, the information derived from the ANN was far less reliable. Firstly, the level of uncertainty for specific variables from Olden's connection weight (OCW) algorithm was very large between model fits for the same network architecture. This suggests that a single model may provide misleading information, and potentially require additional models to reduce uncertainty.

Secondly, when comparing the results from influence scores against OCW, it becomes apparent that the OCW makes very strong assumptions about relationships between predictor and target variables. For example, the influence score of limpets in our barnacle model showed a non-monotonic effect. On the other hand, the ANN connection weights showed a strong positive effect. From an ecological standpoint, limpets have been documented to reside in both macroalgae and barnacle stands (Thompson, 1980). However, limpets can also graze and prevent the settling of barnacle larvae, which would also lead to a negative relationship between the two variables (Blackmore, 1969). In this instance, the BN is correct in attributing limpets having a non-monotonic effect on barnacles. This also suggests that the strong assumptions made by OCW can provide misleading information about variable importance, a problem that may be exacerbated when applied to more complex ecological systems.

5.5 ANN significantly outperformed GLM models

MLRs and GLMs are the most frequently used predictive tool in ecology largely due to their ease of use and its ability to give explanatory results, as the regression coefficients of input variables provide simple interpretable information about their relative importance (Laë et al., 1999). However, MLR models make strong assumptions about distributions of the data, therefore they are unable to handle non-linear relationships between dependent and independent variables. While GLMs can be used to handle non-linear effects, ANNs showed superior predictive capabilities, while making

minimal assumptions about the underlying distribution of the data. This was clearly reflected in our results, where the ANNs performed significantly better than GLM.

5.6 General conclusion

To our knowledge, this is the first application of a hybrid style Bayesian network-Neural network modelling approach in the study of ecology. BN algorithms were able to recover known functional links of a typical rocky shore community, while providing insights to relative variable importance through Markov blankets. This was then utilized as a novel variable selection method to train an artificial neural network on each component species of the revealed network. The results from this workflow showed general improvements in predictive performance in models with variable selection versus no variable selection. The exception in grazers highlighted the need for physical factors to be considered as well. Additionally, ANN significantly outperformed conventional GLM models. It should be noted that these models were applied to a relatively simple system, one with very few variables. Therefore, this difference in performance may be even more pronounced when applied to a more complex system. Further, variable selection in a more complex system could enable future data collection to focus on those features relevant to those variables whose prediction is of interest, potentially reducing experimental effort.

In conclusion, Bayesian networks show strong potential in revealing ecological networks, however due to discretization of data, it is sub-optimal as a predictive tool as we would be limited to a classification type problem. On the other hand, artificial neural networks provide an opportunity to overcome this problem as it can handle complex non-linear data, while showing strong predictive capabilities. However, despite the ease of use and its high predictive abilities, a commonly cited weakness of ANN is its black box nature, where the results are generally difficult if not impossible to interpret. Given that explaining ecological relationships is integral in the study of ecology, ANNs may not be optimal as a standalone model. Here, we demonstrate that coupling the results from the Bayesian network can complement the strong predictive abilities of ANNs, as relative importance and contributions of each variable can be recovered from the BN. This in turn can overcome each models' respective shortcomings. Therefore, this Bayesian network-neural network approach shows promise in the field of ecology as it can achieve two important features: 1) it has potential to reveal ecological network structures in different ecosystems, where existing relationships between species and other functional components are not known; and 2) it can guide the training of powerful predictive models by serving as a robust variable selection tool.

Acknowledgements

Funding: This work was supported by St Leonard's Postgraduate College of the University of St Andrews.

References

- Arrontes, J., Arenas, F., Fernández, C., Rico, J., Oliveros, J., Martínez, B., Viejo, R.M., Alvarez, D., 2004. Effect of grazing by limpets on mid-shore species assemblages in northern Spain. *Mar. Ecol. Prog. Ser.* 277, 117-133. <https://doi.org/10.3354/meps277117>
- Beck, M.W., 2018. NeuralNetTools: Visualization and analysis tools for neural networks. *J. Stat. Softw.* 85, 1–20. <https://doi.org/10.18637/jss.v085.i11>
- Blackmore, D.T., 1969. Studies of *Patella vulgata* L. I. Growth, reproduction and zonal distribution. *J. Exp. Mar. Bio. Ecol.* 3, 200–213. [https://doi.org/10.1016/0022-0981\(69\)90018-5](https://doi.org/10.1016/0022-0981(69)90018-5)
- Brosse, S., Guegan, J.F., Tourenq, J.N., Lek, S., 1999. The use of artificial neural networks to assess fish abundance and spatial occupancy in the littoral zone of a mesotrophic lake. *Ecol. Modell.* 120, 299–311. [https://doi.org/10.1016/S0304-3800\(99\)00110-6](https://doi.org/10.1016/S0304-3800(99)00110-6)
- Catesby, S.M., McKillup, S.C., 1998. The importance of crevices to the intertidal snail *Littoraria articulata* (Philippi) in a tropical mangrove forest. *Hydrobiologia* 367, 131–138. <https://doi.org/10.1023/A:1003271915241>
- Chapman, M.G., Underwood, A.J., 1994. Dispersal of the intertidal snail, *Nodilittorina pyramidalis*, in response to the topographic complexity of the substratum. *J. Exp. Mar. Bio. Ecol.* 179, 145–169. [https://doi.org/10.1016/0022-0981\(94\)90111-2](https://doi.org/10.1016/0022-0981(94)90111-2)
- Chen, S., Mar, J.C., 2018. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics* 19, 1–21. <https://doi.org/10.1186/s12859-018-2217-z>
- Davies, A. J., Johnson, M. P., Maggs, C.A., 2007. Limpet grazing and loss of *Ascophyllum nodosum* canopies on decadal time scales. *Mar. Ecol. Prog. Ser.*, 339, 131–141. <https://doi.org/10.3354/meps339131>
- Folke, C., Carpenter, S., Walker, B., Scheffer, M., Elmqvist, T., Gunderson, L., Holling, C.S., 2004.

- Regime shifts, resilience, and biodiversity in ecosystem management. *Annu. Rev. Ecol. Evol. Syst.* 35, 557–581. <https://doi.org/10.1146/annurev.ecolsys.35.021103.105711>
- Friedman, N., Linial, M., Nachman, I., Pe'er, D., 2000. Using Bayesian Networks to Analyze Expression Data. *J. Comput. Biol.* 7, 601–620. <https://doi.org/10.1089/106652700750050961>
- Fritsch, S., Guenther, F., Wright, M.N., 2019. neuralnet: Training of Neural Networks. R package version 1.44.2. <https://CRAN.R-project.org/package=neuralnet>
- Gevrey, M., Dimopoulos, I., Lek, S., 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecol. Modell.* 160, 249–264. [https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0)
- Hausser, J., Strimmer, K., 2009. Entropy inference and the James-Stein estimator, with application to nonlinear gene association networks. *J. Mach. Learn. Res.* 10, 1469–1484.
- Hawkins, S.J., Bohn, K., Doncaster, C.P., 2015. Ecosystems: The Rocky Road to Regime-Shift Indicators. *Curr. Biol.* 25, R666–R669. <https://doi.org/10.1016/j.cub.2015.06.027>
- Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., Guthke, R., 2009. Gene regulatory network inference: Data integration in dynamic models—A review. *Biosystems* 96, 86–103. <https://doi.org/10.1016/j.biosystems.2008.12.004>
- Henneman, M.L., Memmott, J., 2001. Infiltration of a Hawaiian Community by Introduced Biological Control Agents. *Science* 293, 11314–1316. <https://doi.org/10.1126/science.1060788>
- Hidalgo, F.J., Firstater, F.N., Fanjul, E., Bazterrica, M.C., Lomovasky, B.J., Tarazona, J., Iribarne, O.O., 2008. Grazing effects of the periwinkle *Echinolittorina peruviana* at a central Peruvian high rocky intertidal. *Helgol. Mar. Res.* 62, 73–83. <https://doi.org/10.1007/s10152-007-0086-3>
- Hui, E., Stafford, R. Matthews, I.M., Smith, V.A., 2021. Rocky Shore Samples: Bayesian Networks as a novel tool to enhance interpretability and predictive power of ecological models. Dataset. University of St Andrews Research Portal. <https://doi.org/10.17630/f2b69f88-efb7-43a1-96e9-70012256a752>
- Jeong, K.-S., Joo, G.-J., Kim, H.-W., Ha, K., Recknagel, F., 2001. Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. *Ecol. Modell.* 146, 115–129. [https://doi.org/10.1016/S0304-3800\(01\)00300-3](https://doi.org/10.1016/S0304-3800(01)00300-3)
- Jerkanof, P., 2006. Interactions between the limpet *Patelloida latistrigata* and algae on an intertidal rock. *Marine Ecol.* 23, 71–78.

- Jonsson P.R., Granhag L., Moschella P.S., Aberg P., Hawkins S.J., Thompson R.C., 2006. Interactions between wave action and grazing control the distribution of intertidal macroalgae. *Ecology* 87, 1169–1178. [https://doi.org/10.1890/0012-9658\(2006\)87\[1169:ibwaag\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[1169:ibwaag]2.0.co;2)
- Kroodsma, D.A., Mayorga, J., Hochberg, T., Miller, N.A., Boerder, K., Ferretti, F., Wilson, A., Bergman, B., White, T.D., Block, B.A., Woods, P., Sullivan, B., Costello, C., Worm, B., 2018. Tracking the global footprint of fisheries. *Science* 359, 904–908. <https://doi.org/10.1126/science.aao5646>
- Kuhn, M., 2020. caret: Classification and Regression Training. R package version 6.0-86. <https://CRAN.R-project.org/package=caret>
- Laë, R., Lek, S., Moreau, J., 1999. Predicting fish yield of African lakes using neural networks. *Ecol. Modell.* 120, 325–335. [https://doi.org/10.1016/S0304-3800\(99\)00112-X](https://doi.org/10.1016/S0304-3800(99)00112-X)
- Lantz, B., 2013. Machine Learning with R. Packt Publishing (Birmingham) ISBN 978-1-78216-214-8.
- Mac Aodha, O., Gibb, R., Barlow, K.E., Browning, E., Firman, M., Freeman, R., Harder, B., Kinsey, L., Mead, G.R., Newson, S.E., Pandourski, I., Parsons, S., Russ, J., Szodoray-Paradi, A., Szodoray-Paradi, F., Tilova, E., Girolami, M., Brostow, G., Jones, K.E., 2018. Bat detective—Deep learning tools for bat acoustic signal detection. *PLOS Comput. Biol.* 14, e1005995. <https://doi.org/10.1371/journal.pcbi.1005995>
- Milns, I., Beale, C.M., Anne Smith, V., 2010. Revealing ecological networks using Bayesian network inference algorithms. *Ecology* 91, 1892–1899. <https://doi.org/10.1890/09-0731.1>
- Mitchell, E.G., Wallace, M.I., Smith, V.A., Wiesenthal, A.A., Brierley, A.S., 2021. Bayesian Network Analysis reveals resilience of the jellyfish *Aurelia aurita* to an Irish Sea regime shift. *Sci. Rep.* 11, 1–14. <https://doi.org/10.1038/s41598-021-82825-w>
- Nicotri, R.E., 1977. Grazing effects of four marine intertidal herbivores on the microflora. *Ecology* 58, 1020–1032.
- Norton, T.A., Hawkins, S.J., Manley, N.L., Williams, G.A., Watson, D.C., 1990. Scraping a living: a review of littorinid grazing. *Hydrobiologia* 193, 117–138. <https://doi.org/10.1007/BF00028071>
- Olden, J.D., Jackson, D.A., 2002. Illuminating the “black box”: Understanding variable contributions in artificial neural networks. *Ecol. Modell.* 154, 135–150.
- Olden, J.D., Joy, M.K., Death, R.G., 2004. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecol. Modell.* 178, 389–397. <https://doi.org/10.1016/j.ecolmodel.2004.03.013>

- Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Glob. Ecol. Biogeogr.* 12, 361–371.
<https://doi.org/10.1046/j.1466-822X.2003.00042.x>
- Perez, L. and Wang, J., 2017. The effectiveness of data augmentation in image classification using deep learning. *CoRR*, abs/1712.04621
- Pereira, T.D., Aldarondo, D.E., Willmore, L., Kislin, M., Wang, S.S.-H., Murthy, M., Shaevitz, J.W., 2019. Fast animal pose estimation using deep neural networks. *Nat. Methods* 16, 117–125.
<https://doi.org/10.1038/s41592-018-0234-5>
- Petraitis, P.S., Dudgeon, S.R., 1999. Experimental Evidence for the Origin of Alternative Communities on Rocky Intertidal Shores. *Oikos* 84, 239–245.
- Petraitis, P.S., Methratta, E.T., Rhile, E.C., Vidargas, N.A., Dudgeon, S.R., 2009. Experimental confirmation of multiple community states in a marine ecosystem. *Oecologia* 161, 139–148.
<https://doi.org/10.1007/s00442-009-1350-9>
- Petraitis, P.S., Rhile, E.C., Dudgeon, S., 2003. Survivorship of juvenile barnacles and mussels: spatial dependence and the origin of alternative communities. *J. Exp. Mar. Bio. Ecol.* 293, 217–236.
[https://doi.org/10.1016/S0022-0981\(03\)00219-3](https://doi.org/10.1016/S0022-0981(03)00219-3)
- Proulx, S.R., Promislow, D.E.L., Phillips, P.C., 2005. Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353. <https://doi.org/10.1016/j.tree.2005.04.004>
- Raffaelli, D., Hawkins, S., 1999. *Intertidal Ecology*. Kluwer Academic Publishers (Dordrecht)
- Scheffer, M., Carpenter, S., Foley, J.A., Folke, C., Walker, B., 2001. Catastrophic shifts in ecosystems. *Nature* 413, 591–596.
- Scheffer, M., Hosper, S.H., Meijer, M.-L., Moss, B., Jeppesen, E., 1993. Alternative equilibria in shallow lakes. *Trends Ecol. Evol.* 8, 275–279. [https://doi.org/10.1016/0169-5347\(93\)90254-M](https://doi.org/10.1016/0169-5347(93)90254-M)
- Seuront, L., Ng, T.P.T., 2016. Standing in the sun: Infrared thermography reveals distinct thermal regulatory behaviours in two tropical high-shore littorinid snails. *J. Molluscan Stud.* 82, 336–340. <https://doi.org/10.1093/mollus/eyv058>
- Smith, V.A., Yu, J., Smulders, T. V., Hartemink, A.J., Jarvis, E.D., 2006. Computational Inference of Neural Information Flow Networks. *PLOS Comput. Biol.* 2, e161.
<https://doi.org/10.1371/journal.pcbi.0020161>
- Stafford, R., 2002. The role of environmental stress and physical and biological interactions on the

ecology of high shore littorinids in a temperate and a tropical region. PhD Thesis, University of Sunderland.

Stafford, R., Ng, T.P.T., Williams, G.A., Davies, M.S., 2015. A biologically relevant rapid quantification of physical and biological stress profiles on rocky shores. *Ecol. Inform.* 25, 43–48.

<https://doi.org/10.1016/j.ecoinf.2014.11.006>

Stafford, R., Smith, V.A., Husmeier, D., Grima, T., Guinn, B., 2013. Predicting ecological regime shift under climate change: New modelling techniques and potential of molecular-based approaches 59, 403–417. <https://doi.org/10.1093/czoolo/59.3.403>

Steffen, W., Rockström, J., Richardson, K., Lenton, T.M., Folke, C., Liverman, D., Summerhayes, C.P., Barnosky, A.D., Cornell, S.E., Crucifix, M., Donges, J.F., Fetzer, I., Lade, S.J., Scheffer, M., Winkelmann, R., Schellnhuber, H.J., 2018. Trajectories of the Earth System in the Anthropocene. *Proc. Natl. Acad. Sci.* 115, 8252–8259.

<https://doi.org/10.1073/pnas.1810141115>

Thompson, G.B., 1980. Distribution and population dynamics of the limpet *Patella vulgata* L. in Bantry Bay. *J. Exp. Mar. Bio. Ecol.* 45, 173–217. [https://doi.org/10.1016/0022-0981\(80\)90058-1](https://doi.org/10.1016/0022-0981(80)90058-1)

Trifonova, N., Kenny, A., Maxwell, D., Duplisea, D., Fernandes, J., Tucker, A., 2015. Spatio-temporal Bayesian network models with latent variables for revealing trophic dynamics and functional networks in fisheries ecology. *Ecol. Inform.* 30, 142–158.

<https://doi.org/10.1016/j.ecoinf.2015.10.003>

van Wijk, M.T., Bouten, W., 1999. Water and carbon fluxes above European coniferous forests modelled with artificial neural networks. *Ecol. Modell.* 120, 181–197.

[https://doi.org/10.1016/S0304-3800\(99\)00101-5](https://doi.org/10.1016/S0304-3800(99)00101-5)

Yu, J., Smith, V.A., Wang, P.P., Hartemink, A.J., Jarvis, E.D., 2004. Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20, 3594–3603. <https://doi.org/10.1093/bioinformatics/bth448>

APPENDIX A

Figure A1: Frequency distribution of percentage cover and counts of species found in our quadrats.

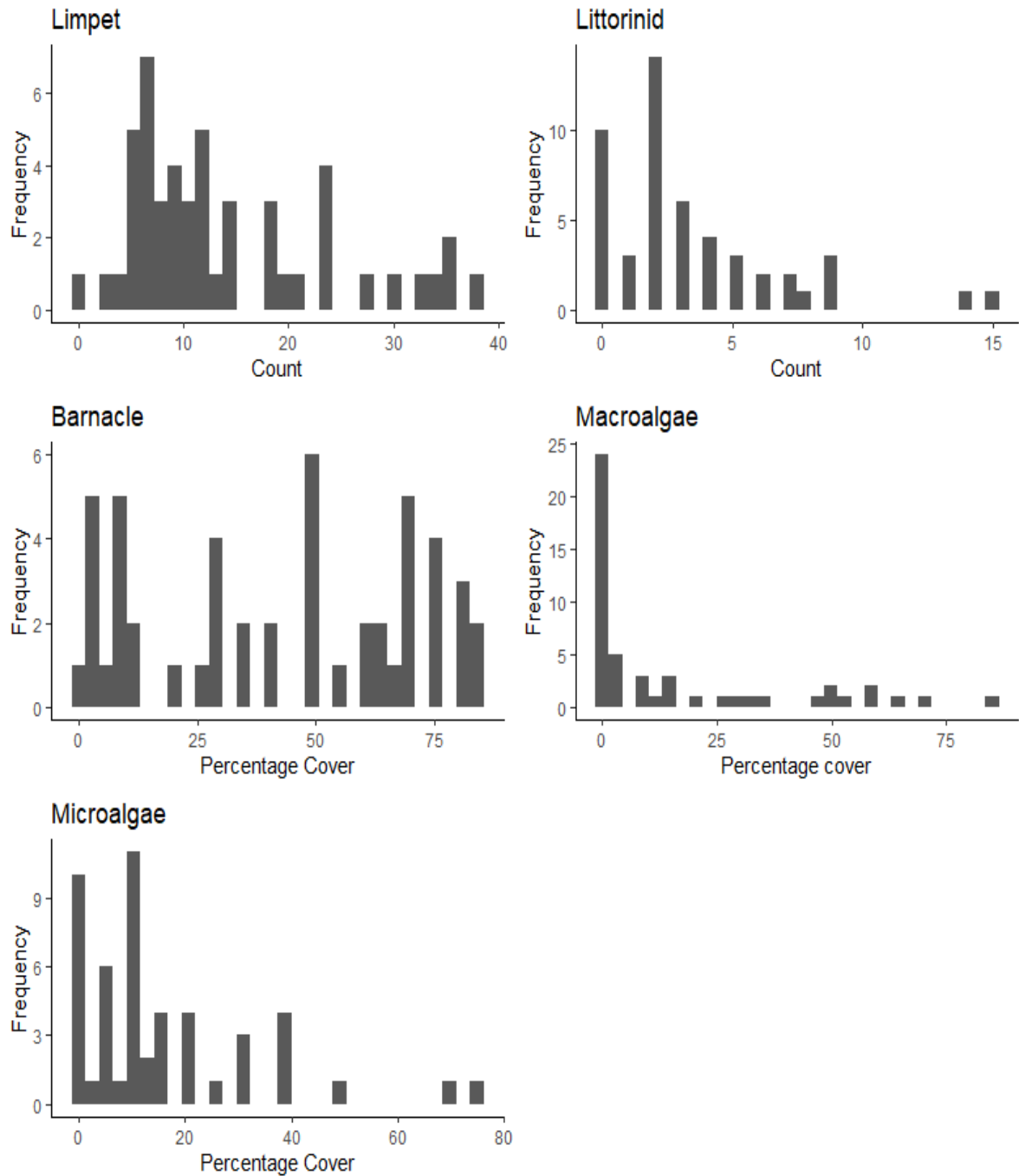


Figure A2: Non-metric dimensional scaling (NMDS) plot of field data.

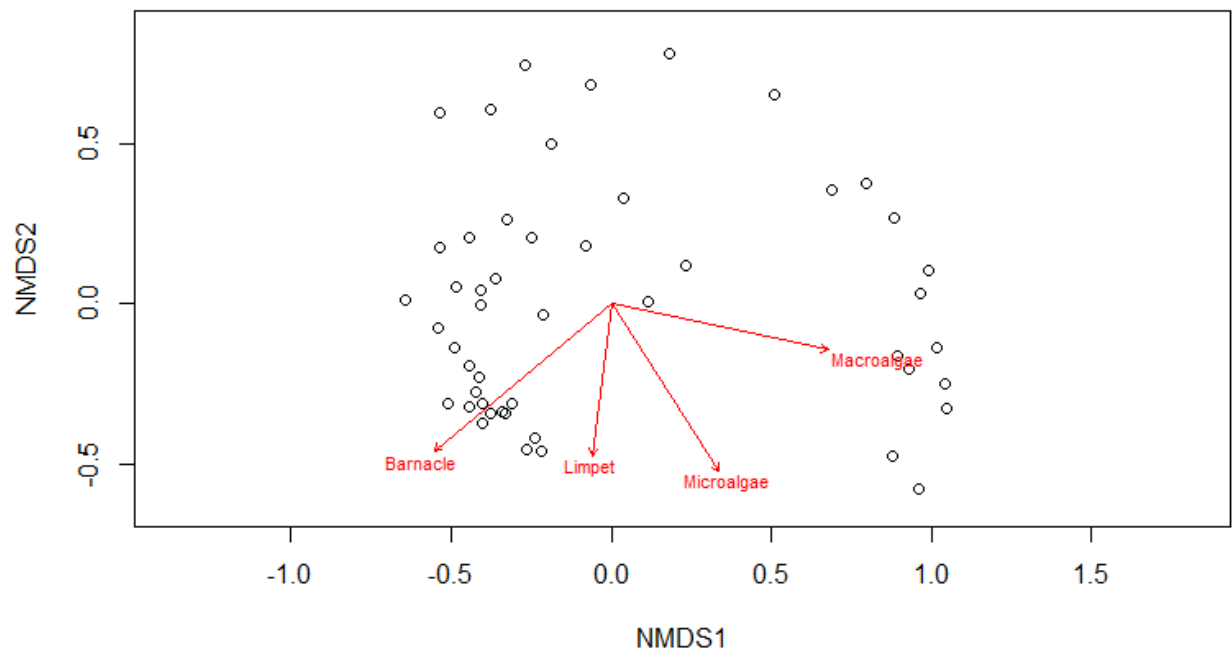


Figure A3: Distribution of field data after quantile discretization into 3 bins.

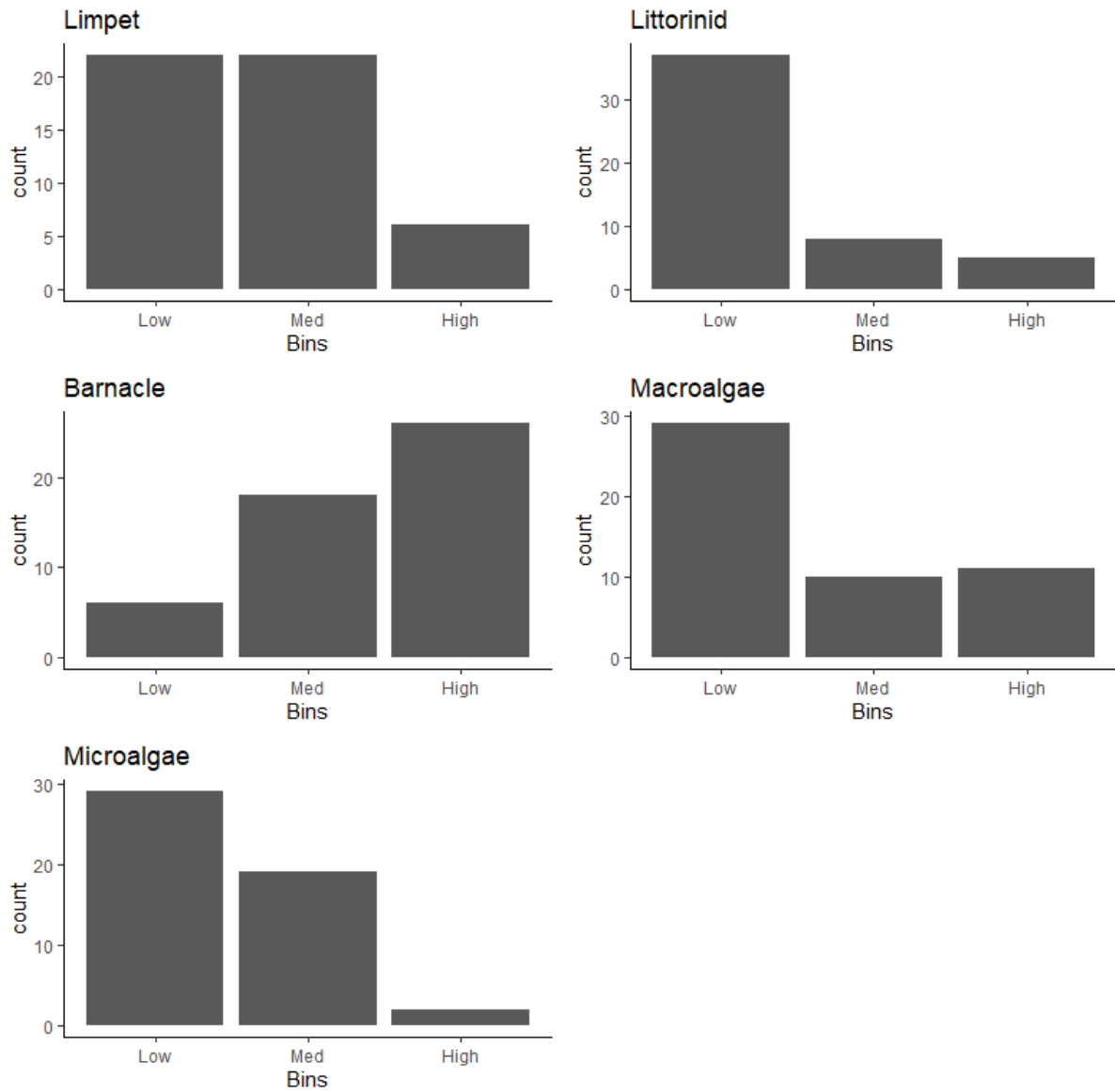


Figure A4: Scatterplot of ANN predicted values versus true values.

