

# Explainable neural networks for trait-based multispecies distribution modelling—A case study with butterflies and moths

Yoann Bourhis<sup>1</sup>  | James R. Bell<sup>1</sup>  | Chris R. Shortall<sup>1</sup>  | William E. Kunin<sup>2</sup>  |  
Alice E. Milne<sup>1</sup> 

<sup>1</sup>Rothamsted Research, Harpenden, UK

<sup>2</sup>Leeds University, Leeds, UK

## Correspondence

Yoann Bourhis

Email: [yoann.bourhis@rothamsted.ac.uk](mailto:yoann.bourhis@rothamsted.ac.uk)

## Funding information

Biotechnology and Biological Sciences Research Council, Grant/Award Number: BBS/E/C/000.J0200; Natural Environment Research Council, Grant/Award Number: NE/V00686X/1

Handling Editor: Natalie Cooper

## Abstract

1. Species response traits mediate environmental effects on species distribution. Traits are used in joint and multispecies distribution models (JSDMs and MSDMs) to enable community-wide shared parameters that characterise niche filtering along environmental gradients. Multispecies machine learning SDMs, however, do not use traits as their inclusion requires an additional taxonomic dimension that is incompatible with their usual tabular inputs. This has confined trait mediation in SDMs to hierarchical Bayesian models. Here we provide a novel artificial neural network (ANN) architecture that solves this dimensionality problem.
2. Our ANN includes species traits (via a time distributed layer) and is therefore able to identify not only species-specific responses to the environment, but also shared responses across the community that are mediated by species traits. Model performance evaluated at the species level not only quantifies the reliability of species predictions, but also their departure from an average response dictated by traits only.
3. We apply our model to two unique long-term spatio-temporal of butterfly and moth datasets collected across the United Kingdom between 1990 and 2019. In addition to species traits, predictors include numerous metrics derived from weather, land-cover and topology data. For butterflies and moths we show convincing model performance for classifying species occupancy. We use SHAP (Shapley Additive exPlanations) to explain the ANN and show how trait-mediated and species-specific responses can be approximated, hence yielding ecological insights on the key drivers of species distribution. We highlight a range of drivers of change that determine occupancy, including wind, temperature as well as habitat type.
4. We demonstrate that a trait-based approach can be encoded as an ANN by using a time distributed layer. This brings ANNs unmatched predictive capabilities to the field of MSDMs, at the same time of lifting their reputed drawback of poor explainability.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Methods in Ecology and Evolution* published by John Wiley & Sons Ltd on behalf of British Ecological Society.

## KEYWORDS

ANN, Lepidoptera, MSDM, SHAP, traits

## 1 | INTRODUCTION

The geographic distribution of species defines local community compositions that result from complex assembly processes. These compositions are shaped by biotic and abiotic filters that can leave predictable signals on species distributions patterns (Münkemüller et al., 2020). Building on those signals, we can investigate the driving forces affecting species distribution along environmental gradients, paving the way to further understanding of global change impacts on observed biodiversity shifts. In this regard, quantifying effects of, for example, intensified land use, climate change, pollution or novel pests and pathogens constitute significant steps towards the preservation of ecosystem services and conservation in general.

The underlying mechanisms of species distribution are ideally explored through mechanistic models, yielding causal understanding of the rules and trends in biodiversity (see e.g. Kempel et al., 2015). The interactive nature of those mechanisms, however, calls for holistic modelling approaches which conflict with the need for keeping parameters identifiable by limiting model complexity and colinear inputs. Although they are data hungry, high-throughput correlative methods are less affected by this limit. They can search through data, filtering for patterns prior to any attempt at deciphering the system mechanistically, and as such have been advocated as powerful tools enabling the subsequent investigation of causal links (Baker et al., 2018).

The huge number of parameters involved in putative community assembly processes is also troublesome for statistical models. Yet, thanks to hierarchical structures, Bayesian approaches that build on Markov chain Monte Carlo (MCMC) methods are able to deal with this complexity and to explain community compositions (Bystrova et al., 2021; Ovaskainen & Abrego, 2020; Tikhonov et al., 2020; Zhang et al., 2020). Still, a significant prior filtering of environmental inputs may be necessary for the MCMC to converge with reasonable computing resources. In regard of the increasing availability of remote sensing environmental data, such prior filtering is not necessarily detrimental as it protects the user against overfitting, yet it demands knowledge for selecting what information to include or not.

Key elements in linking environment and species distribution are species traits. Traits are discrete or continuous descriptors of species. Response traits mediate niche filtering along environmental gradients, which drives groups of species into community compositions. Joint and multispecies distribution models (JSDMs and MSDMs) can explicitly include these species-level, community-wide response traits to predict species occurrence (or abundance) and elucidate functional insights (Ovaskainen et al., 2017; Pollock et al., 2012). These insights take the form of an interaction coefficient between an environmental covariate and a species trait, for example, a negative relationship between an insect's body size and altitude would suggest that larger species are less suited for mountains. Highlighting trait–environment relationships with

such a correlative approach can give cues on underlying mechanisms and serve in the making of competing mechanistic models to investigate causal links of species distribution. It can also help achieving better predictions in poorly recorded species, providing their traits are known.

Here we describe an artificial neural network (ANN) architecture able to tackle the problem of species assembly into communities by building on trait mediation and multispecies distribution modelling. Building on TensorFlow and being trained through backpropagation instead of MCMC, it is less affected by the computational limits exhibited in other trait-based approaches (e.g. *Hmsc*). It is therefore particularly suited for an agnostic approach in which inputs (species traits and environmental covariates) may be numerous. Previous machine learning (ML) approaches have first addressed community composition through stacked single-output (i.e. species specific) SDMs (Calabrese et al., 2014; Guisan & Rahbek, 2011; Williams et al., 2009). Recent ML developments have for example targeted more explicit spatial predictions with convolutional layers (Deneu et al., 2021), or the learning and prediction of joint distributions (as in JSDMs) that account for species interactions (i.e. biotic filtering, Harris, 2015; Pichler & Hartig, 2021). The novelty in our approach is the identification of trait-mediated shared responses. Those are responses to environmental gradients that are mediated by one or more traits, and hence apply to all species as a function of their trait value. In practice, our approach builds on two components: the first component learns the shared responses through parameters that are common across species, while the second component is trained to learn species-specific responses and allow for further flexibility. Then, using SHAP (SHapley Additive exPlanation, Lundberg & Lee, 2017), we go beyond the usual 'black box' and quantify the role of the different inputs (species traits and environmental covariates) in the model predictions. The environment filtering is rendered in a way that makes ecological sense: through a matrix of trait-mediated shared responses (i.e. the fourth corner matrix Dolédec et al., 1996; Legendre et al., 1997), and a matrix of species-specific niches. As a case study, our developments are applied to butterfly and moth occurrence data recorded across the United Kingdom between 1990 and 2019.

## 2 | MATERIALS AND METHODS

### 2.1 | Data

#### 2.1.1 | Community

We explore two sets of insect community data. The first is a dataset from the UK Butterfly Monitoring Scheme (UKBMS) consisting of 59 butterfly species collected across the United Kingdom between 1990 and 2019 (Brereton et al., 2020). Its measurements take the form of yearly site indices ( $n=25,000$ ) quantifying the local abundance of a

species for a given year (Dennis et al., 2016). The second dataset comprises yearly counts of 461 species (selected for having at least 20 detection events) of macromoths ( $n = 2300$ ) from the Rothamsted Insect Survey light trap network (Bell et al., 2020; Fox et al., 2020; Storkey et al., 2016), collected across Great Britain between 1990 and 2018. By their dimensions ( $59 \times 25,000$  and  $461 \times 2300$ ), the two datasets constitute contrasting challenges to learning (although they aggregate about the same amount of information). The following developments apply on a binarised version of those datasets, as we focus on the prediction of species occupancy (aka presence/absence).

## 2.1.2 | Environment

The environment data come from several sources. First, the BIOCLIM19 set of climatic variables are derived from monthly temperature averages, minima and maxima, as well as monthly rainfall (sourced from the Had UK 1 km  $\times$  1 km grid Hollis et al., 2019). BIOCLIM19 variables capture trends, seasonality and extrema that potentially affect the organisms of interest, and are therefore common inputs in SDM (see e.g. Hill et al., 2017). Second, the UK CEH Land-cover map (1990, 2000, 2007, 2015, 2017, 2018 and 2019 editions; Morton et al., 2020) defines 10 aggregated land-cover types across the United Kingdom at a 25 m  $\times$  25 m resolution. We aggregate those to our working 1 km  $\times$  1 km resolution, deriving composition and diversity metrics to capture landscape complexity. The remaining variables are static: these are the river network density (from UKCEH and DAERA), terrain (elevation, slope and aspect from AWS terrain tiles) and distance to sea. See Table A1 in Appendix A for the details of the environmental covariates.

## 2.1.3 | Traits

The butterfly's traits come from two trait databases (Cook et al., 2021; Middleton-Welling et al., 2020) from which were selected traits that were fully informed for the 59 species of butterflies encountered in the United Kingdom. In total seven traits were retained: wing index, wing index variation, voltinism, overwintering stage, number of host plants, host plant types and number of habitat types. From the moth trait database (Cook et al., 2021), six traits similar to the butterfly ones were also fully informed: forewing length, forewing variation coefficient, voltinism, overwintering stage, number of host plants and number of habitats. Additionally, for each community, the taxonomic tree of the composing species was vectorised and added as a supplementary trait to possibly capture a phylogenetic signal (as in Ovaskainen & Abrego, 2020).

## 2.2 | Model architecture

The model is composed of two components: one learning the community's shared responses to the environment, and one learning the species-specific responses. Both components produce, for every sample (i.e. every site  $\times$  year), a tabular output composed of the probability of

occurrence of the numerous species of interest (i.e. a table with dimension  $n$  samples  $\times$   $q$  species). Such a multi-output approach implies shared capacity within the hidden layers that are common to all species. This allows shared representation to be learnt prior to the output layer, hence enabling species to build on one another during training and achieve greater performance. The two components take different inputs:

1. *Shared responses*. For a given sample, the only information relevant to all species is the environmental covariates. Hence to input the traits' values we build on a Kronecker product of the traits and the environmental covariates (i.e. every trait is multiplied to every covariate). The traits being informed for every species, that input has three dimensions:  $n$  samples  $\times$   $m$  covariates  $\cdot$  ( $p$  traits + 1 intercept)  $\times$   $q$  species. Note that the intercept allows for the identification of a nonmediated but still shared response of the community as a whole to environmental drivers. In practice, this expands the input row into a matrix of the product values for every species. The use for the Kronecker product may be questioned in the light of the universal approximation theorem (Hornik et al., 1989), as we should expect the relevant trait-environment interactions to be learnt during training, in theory. In practice, however, we found the product to help performance wise and, more importantly, to enable the explanation of the interactions (see Appendix A for details).
2. *Species-specific responses*. To account for the species-specific direct effect of the environmental covariates, the second input is a much more conventional 2D table with the covariate values at each sample point (with dimension  $n$  samples  $\times$   $m$  covariates).

### 2.2.1 | The time distributed layer

The first branch has a 2D input for every sample. It could be processed as such but most inputs are then irrelevant to a given species, only the interactions with its own trait values matters. To maintain a multi-output configuration (with the aforementioned benefits), while ensuring that only relevant inputs make their way across the branch towards a species prediction, a time distributed (TD) layer can be used. This construction is sometimes called a wrapper as it encapsulates layers, in our case simple dense layers. TD layers are used in recurrent ANNs to apply the same weights and biases to every time step of a sequence. In our case, the "time" dimension is the taxonomic dimension, that is, the different species. Here the encapsulated layers will be fitted the same weights and biases, but because the input changes from species to species as a result of changing trait values, the outputs remain species specific. This effectively enforces trait mediation to be learnt.

### 2.2.2 | Merging the two components

The shared and species-specific responses can come together either as two branches of a multi-branch model, or as two models of an

ensemble. These are distinct approaches that implies different training processes. In the multi-branch model, the two branches are trained sequentially. The shared-response branch is trained first and then the species-specific branch is unfrozen (while the former one is frozen, i.e. weights are fixed) to learn species-specific responses which allows further flexibility. In this approach the species-specific responses are trained on the residuals of the shared responses and are therefore conditional to them. Alternatively, in the ensemble approach, both models are trained simultaneously before seeing their outputs averaged. This approach is more flexible as the species-specific responses are not conditional to the shared responses. Figure 1 illustrates the network architecture and dimensions. We advocate here for both options and refer the reader to Appendix A for further details.

## 2.3 | Model training

As often in biology, special care is to be given to class imbalance (Saito & Rehmsmeier, 2015). Some rare species can be easily overlooked by a *lazy* classifier whose trivial predictions of, for example, predicting only absences disregarding the input would score high in accuracy metrics by default. The same reasoning applies to near-ubiquitous species. Therefore, species-specific square root class weights are included into the binary cross-entropy loss function. The weights give more importance to subsamples (i.e. species  $\times$  sample level weights) that includes rarer cases (i.e. the absence of a common species, or the presence of a rare one).

For both butterflies and moths, half the samples were reserved as testing data to assess the models' performances. Among the remaining half, the training dataset, a third was used as validation data to track the learning curve and halt training as soon as overfitting was detected (i.e. after 5 successive epochs without improvement on the validation dataset). For its known robustness to class imbalance, the Matthews correlation coefficient (MCC) was used as a metric in model selection (Chicco & Jurman, 2020; Lever et al., 2016; Saito & Rehmsmeier, 2015). The MCC is a measure of association of two binary variables, which here are the observed occupancy and the binarised predicted probability of

presence. Other relevant metrics are commonly found in the literature, for example, the area under the curve of the receiver operating characteristic (AUC ROC) or precision–recall (PR) curve (Fernández, 2018), but we argue that they are over-optimistic metrics in our case, the former because of strong class imbalance in our data and the later because that imbalance sometimes results of a positive class majority (although being suited for rare species, i.e. a negative class majority, PR and F1 fail to sanction the poor predictions of near-ubiquitous species).

## 2.4 | Model explanation: SHAP

SHAP (Lundberg & Lee, 2017) quantifies the contribution of each feature (or input) to a specific output. It works at the sample scale and therefore builds local explanations whose aggregation to a sufficiently large number of samples can provide model-wide insights. Here, our features are of two types: (1) the products of the trait values with the environmental covariates and (2) the environmental covariates. The procedure calculates a SHAP value for each feature, its contribution, whose averaging over many samples gives to the feature a measure of variable importance to the model outputs (Molnar, 2021). That measure can be positive or negative depending on how the output is affected positively or negatively by the focal feature value, in comparison to a baseline in which the feature is 'deactivated' (i.e. has its sample value replaced with random background values).

Because the procedure builds a collection of explanations made at the sample scale, for every feature we have a collection of SHAP values to linearly regress against a collection of feature values. Hence, the sign of that regression coefficient summarises the effect of the focal feature on the species probability of occurrence (our output).

Here we suggest simplifying the SHAP explanation to a collection of regression coefficients and aggregating them into two figures. First, the effects of the 3D input (i.e. input of the trait-mediated branch), being the same for every species (thanks to the TD layer), it can be reshaped as a  $m$  covariates  $\times$  ( $p$  traits + 1 intercept) matrix, known as fourth corner matrix in ecology (Legendre et al., 1997).

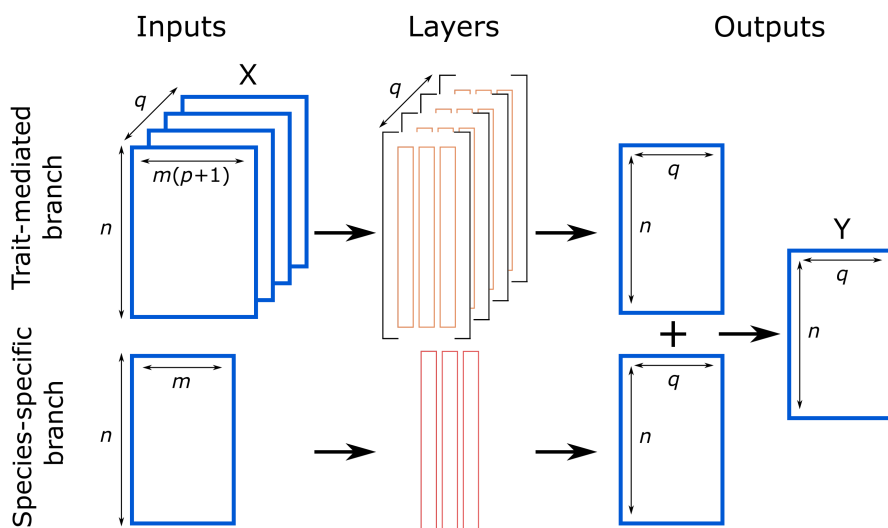


FIGURE 1 Schemata of the ANN architecture. In blue are the tabular inputs ( $X$ ) and outputs ( $Y$ ), in orange and red are the inner layers of the network, the square brackets represent the time distributed layer. The network conforms to the following dimensions:  $q$  species,  $m$  covariates and  $p$  traits.



Second, the effects of the 2D input (i.e. input of the species-specific branch) are collated into a matrix of species-specific effects with shape  $q$  species  $\times$   $m$  covariates.

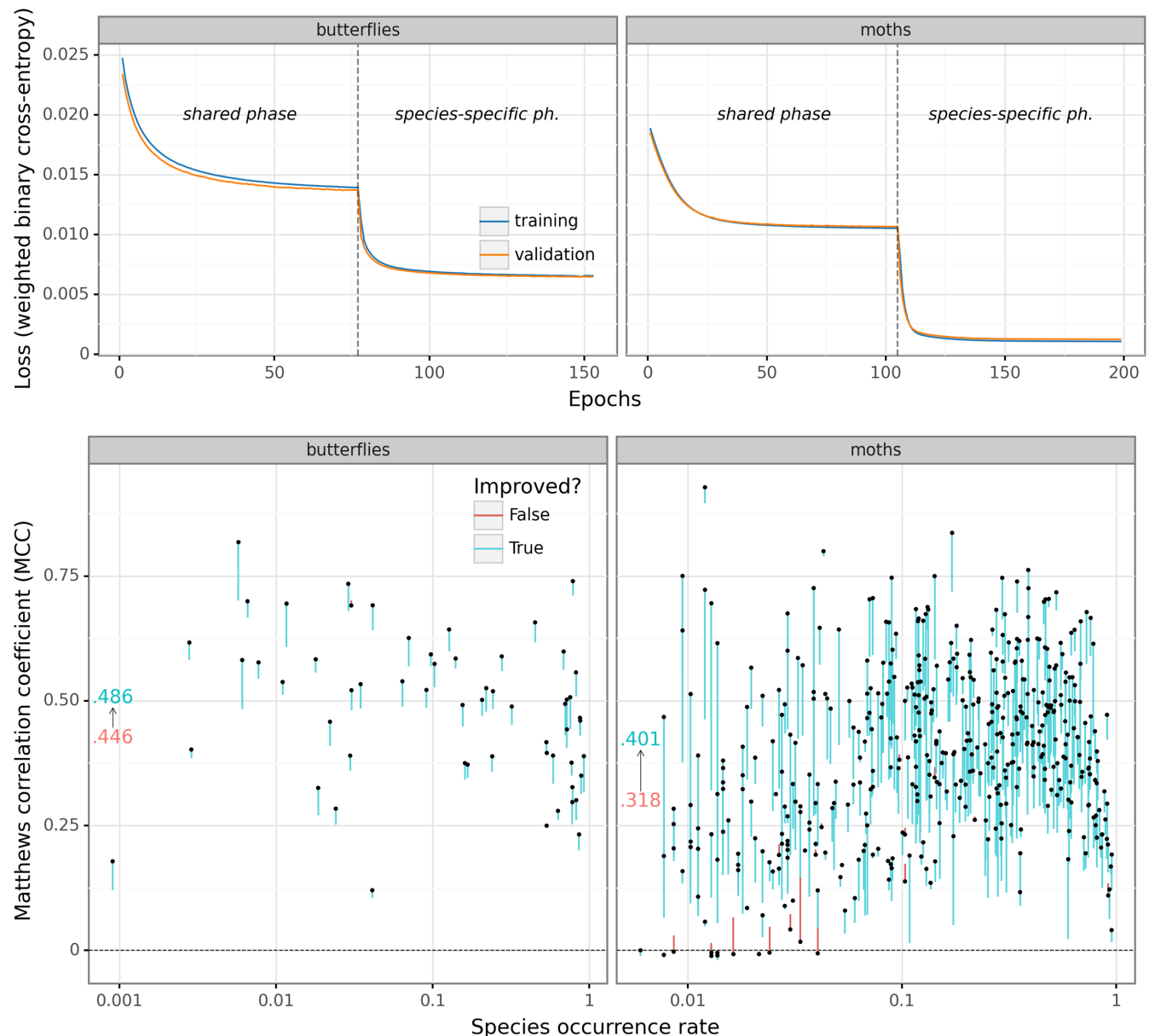
### 3 | RESULTS

#### 3.1 | Training

The learning curves in the top row of Figure 2 illustrates the sequential training of our multi-branch model. The majority of the learning appears to occur in the first phase, in which trait-mediated

shared responses are learnt. Then, after a plateau, as the second branch is unfrozen (i.e. allowed to train weights and biases) and the first one is frozen, further learning occurs. Both training phases are terminated as soon as overfitting is detected. On the other hand, the ensemble having its two models trained simultaneously, its learning curves (not shown) do not feature phases like the multi-branch model. However, the ensemble's performance being marginally better (see Appendix A), the results shown hereafter are derived from it.

Performance wise, according to the AUC ROC that scales from 0 to 1, our models score 0.95 for the butterflies and 0.89 for the moths (all reported metrics are computed on the test dataset). The



**FIGURE 2** Training curves (top row) shows multi-branch models two-step training dynamic and performance following a weighted binary cross-entropy loss function. On the bottom row is the species classifier (ensemble) performances according to the MCC, with the black dot marking the full model performance (shared + species-specific), and the other side of every segment marking the performance of the trait-mediated shared-response only. Most species performances are improved after allowing the additional flexibility of learning species-specific responses.

PR scores are 0.86 and 0.73 (also scaled from 0 to 1). The MCC scales from  $-1$  to  $1$ , with  $0$  marking the random predictions of a *no-skill* classifier. According to this metric, our classifiers score 0.49 and 0.40, but substantial variations are observed from species to species (Figure 2, bottom row). Even if some moth species are predicted no better than at random (especially some of the very rare species), the vast majority of them constitute *skilled predictions*, as is the case for all butterfly species.

The bottom row of Figure 2 also illustrates the performance yielded by each of the two training phases. As expected, most species are better predicted once allowed species-specific responses, still significant performances are already reached after learning only the trait-mediated shared responses. This is especially the case for butterflies in which the trait-mediated shared responses constitute the core learning (short segments with high origins). However, for the moths, species-specific responses are more important to prediction performance (long segments with low origins).

### 3.2 | Prediction

For memory usage concerns, the grid of environmental covariates of the United Kingdom for year 2020 must undergo the Kronecker product and the subsequent model predictions iteratively. It is nonetheless a very inexpensive process from which high-resolution maps can be rendered within seconds. Figure 3 shows the butterfly models predictions for year 2020 across the United Kingdom. There is strong qualitative agreement with known distribution of the 59 butterfly species (<https://www.ukbutterflies.co.uk/distributions.php>). The predicted distributions are highly diverse across species. They mirror no single covariate base layer, making use of the nonlinearities and interactions offered by the model architecture. The MCC score for each species is listed by the species name as a measure of the prediction reliability. The predictions for the 59 most common species of moths are shown in Figure 10 of Appendix C.

Looking at the butterflies, we note that high-performing species include habitat-restricted species such as the rare High Brown Fritillary (*A. adippe*, MCC=0.73) and Swallowtail (*P. machaon*, MCC=0.82) and the more common Gatekeeper (*P. tithonus*, MCC=0.74) and Marbled White (*M. galathea*, MCC=0.64), suggesting that the model captures the drivers that define habitat type well enough to predict the occurrence of specialist butterflies. Conversely, poorly performing species (MCC<0.2) include the Large Blue (*M. arion*, MCC=0.18) and the White-letter Hairstreak (*S. w-album*, MCC=0.12). Both species have suffered major declines in the 20th century through the loss of habitat and host plant (Elm, *Ulmus procera*) respectively. The Large Blue was re-introduced to the United Kingdom in 1983 at certain selected sites that may not correspond to the predicted distribution. The White-letter Hairstreak has recovered from the loss of its host through the Dutch Elm Disease (*Ophiostoma* spp.) which drove declines of the 1970s into areas where Elm suckers and Wych Elm (*Ulmus glabra*) thrive (Thomas & Lewington, 2019); these plant species

distributions may not be well captured by the environmental drivers selected here.

### 3.3 | Explanation

Figure 4 illustrates the butterflies shared responses to the environmental covariates, either mediated by a trait, or unmediated (row *intercept*) but common to all species of interest. Figure 5 summarises the butterflies species niches, that are derived by summing a species trait-mediated shared and species-specific responses for all environmental covariates. Similarly, Figure 6 shows the moths' fourth corner matrix (see Figure 11 in Appendix D for their resulting niches).

In those matrices, the cell colours illustrate the slope of the regression of an input's SHAP value (i.e. importance) to the input value itself. The 1%, 5% and 10% most important inputs, that is, with highest absolute value of feature importance, are highlighted to simplify the interpretation of the matrices. Note that some inputs may be important but still not effectively captured by a simple linear slope (i.e. appearing framed with white background in Figures 4–6 and Figure 11 in Appendix D).

A strong highlight, for example, is the effect of wind, mediated by most butterfly and moth traits. The same observation can be made about the proportion of broadleaf woodlands. Examining the *intercept* shared effects that describe unmediated effects that are nonetheless common to most species, we note the negative effect of wind and positive effect of broadleaf woodlands for both moths and butterflies. For the moth niches (Figure 11 in Appendix D) we observe largely negative impacts of urban area and positive impacts of broadleaf woodlands.

Beyond explaining the model, variable importance also offers a principled way to simplify it by dropping unnecessary inputs. We refer the reader to Appendix A to see how reducing the inputs to the highlighted traits and covariates (in Figures 4 and 6) can affect the models. But to summarise it, we show that our ANNs are barely affected by a drastic reduction of input size, demonstrating how reliable SHAP is for selecting key inputs.

## 4 | DISCUSSION

We demonstrate the use of a time distributed layer as a simple yet solid solution to account for traits in an ANN. This feature brings ANNs unmatched learning abilities to MSDMs, enabling nonlinear and interactive behaviours in a field otherwise dominated by generalised linear models. The TD layer allows for the identification a community's trait-mediated shared responses to environmental covariates. By identifying such functional responses at the community level, species with poor distribution signals (i.e. rare and near-ubiquitous species) can 'borrow strength' (Pollock et al., 2012) from species with stronger definitions, hence producing improved predictions that can then be explained by SHAP. For example, in our case study, wind-trait interactions had near global importance in

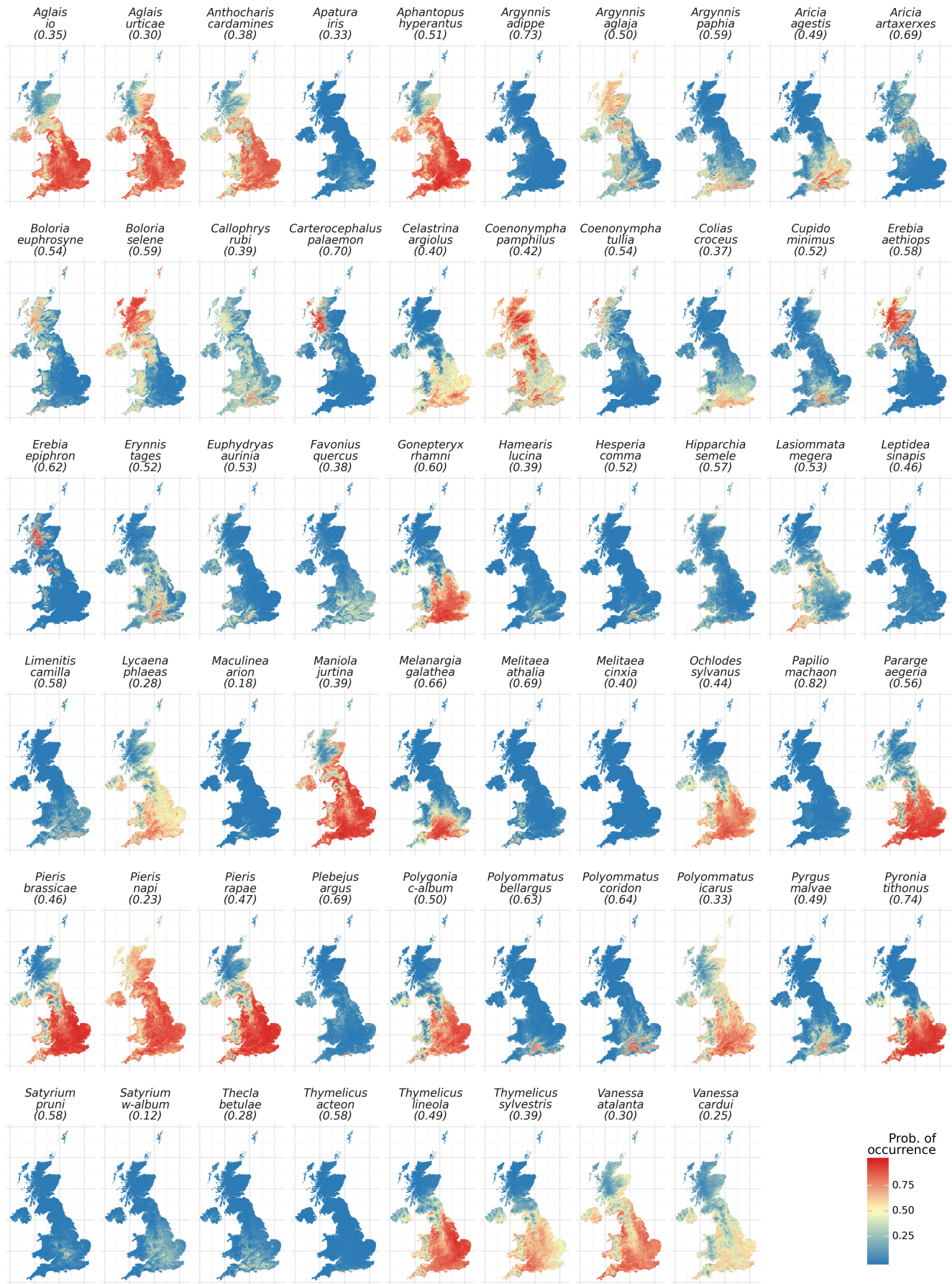
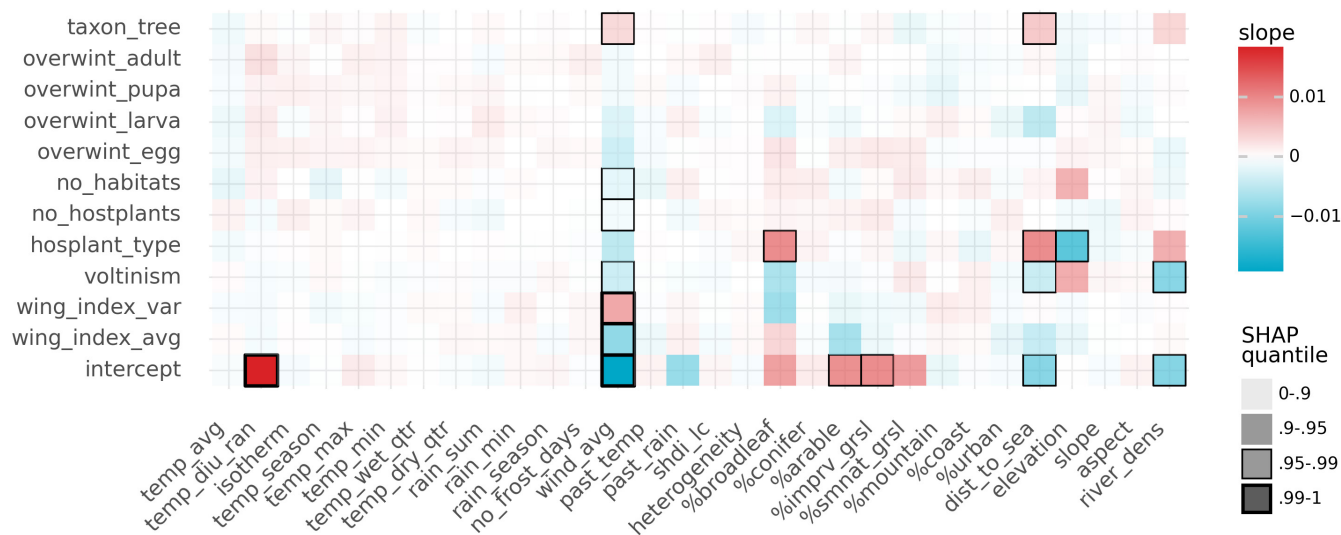


FIGURE 3 Predicted probability of occurrence of the 59 species of butterfly for the year 2020.



**FIGURE 4** Butterflies' trait-mediated shared responses to the environmental covariates. The 1%, 5% and 10% most important inputs are highlighted with darker borders and shades. See covariates meanings in Appendix A.

determining the probability of occurrence for UK butterflies and, to a lesser extent, moths.

Given sufficient width and depth (i.e. neurons and layers), ANNs can compute any function; they are universal approximators (Hornik et al., 1989; Nielsen, 2015). Ecological complexity is no exception, and here our function of computing species distribution from traits and environmental inputs should be within the reach of any sufficiently complex ANN. However, this ability is conditioned on the relevant predictors being among the set of model inputs, which in ecology is a major issue considering the richness of possible inputs. It is also conditioned on the learning algorithm finding the optimal parameters, which is made more difficult as the number of inputs is increased. Here is therefore a dilemma which, if we persist in using numerous unfiltered inputs to pursue an agnostic modelling approach, can be a major obstacle to learning. In this perspective, a trait-based approach is not only an ecologically meaningful depiction of environmental filtering, it is also a solution to the aforementioned dilemma that pools parameters across species and drastically reduces the parameter space. Here, the time distributed layer is the mathematical support of this pooling.

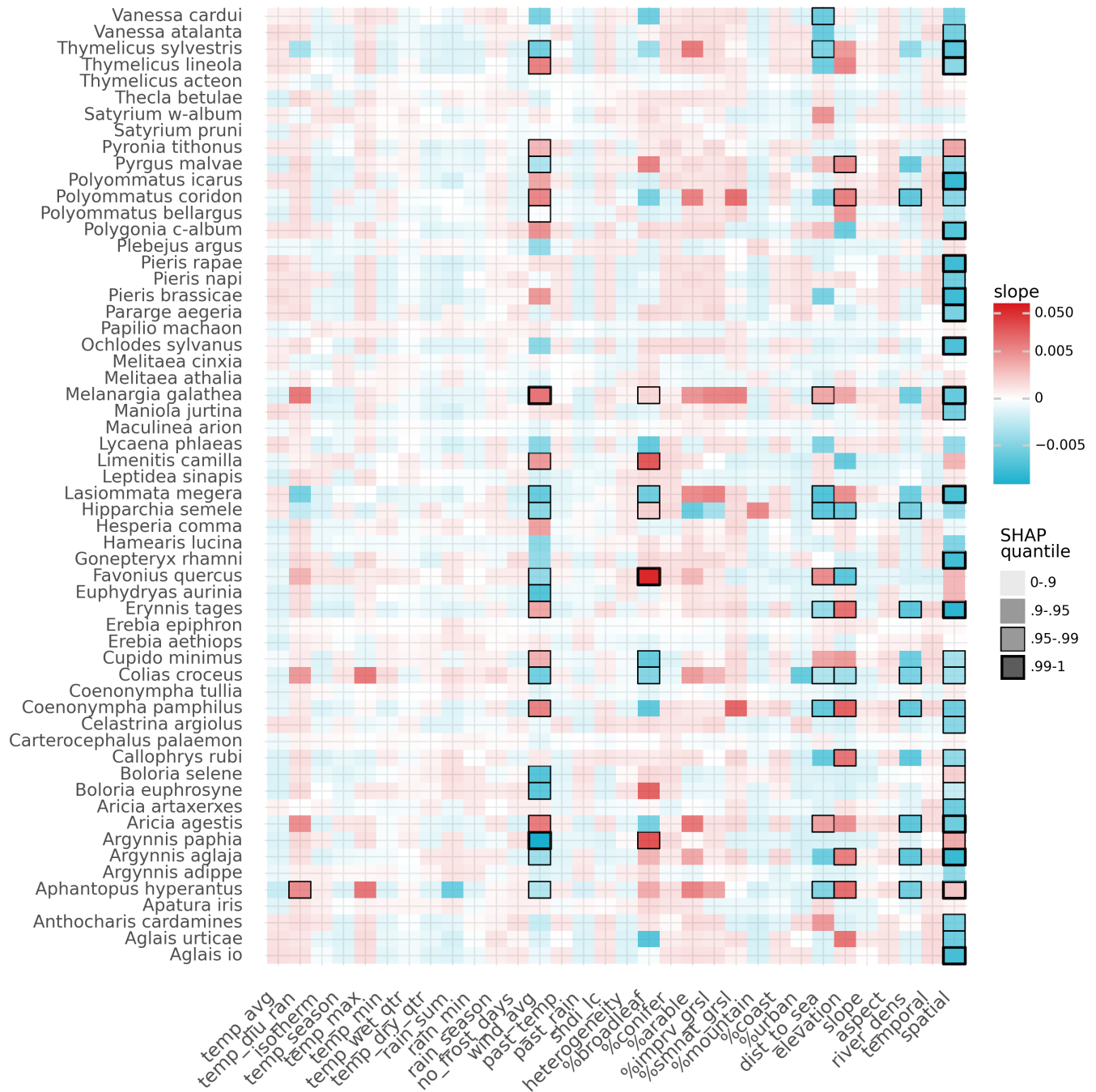
Our models are binary classifiers whose performances are best evaluated with the MCC, a metric avoiding the over-optimism caused by unbalanced classes. According to this metric, butterfly and moth models differ substantially. By comparing the performances reached with the shared response only to the performances of the full model, we can assess how good is the set of traits at explaining the community's occupancy. We see that the butterfly model is only marginally improved by the learning of species-specific responses, therefore we can say that the set of traits used here is of great use to the capture of the shared responses. On the other hand, the opposite is observed for the moth model in which the added flexibility of the species-specific responses is essential to its performance. A possible explanation to this can be found in Figure 9 in Appendix B, which shows how

well the butterfly species scatter in the plane of their trait principal components, showing how well the traits can discriminate between butterfly species. We also see that moth species are not so well discriminated by the selected traits, suggesting that there are few options to shape nonoverlapping niches for those species with the current set of traits.

In SDMs that build on linear models, every parameter carries an ecological meaning, therefore explaining the model is straightforward: the sign of a significant parameter characterises a directed effect of a variable on a species presence or abundance (Dray et al., 2014; Pollock et al., 2012). With ANNs, the parameters are the weights and biases of the successive layers, and none of them per se carry any ecological meaning. The directed effects are not encoded as readily accessible parameters but they are emerging constructs of the network. This strength of ANNs, allowing the capture of nonlinear and interacting behaviours, is also a weakness when it comes to model explanation. As we have seen, the black box can nonetheless be resolved using the SHAP method. By evaluating variable importance at the sample scale, the input's effects on the outputs (i.e. the Shapley values) can be regressed against the input values (i.e. the feature values), and hence characterised at the model scale. We used linear regression for this purpose, and aggregated the slope coefficients into two matrices. In contrast to usual fourth corner matrices which aggregates a linear models' real parameters, ours is only a linear approximation of an ANN's behaviours. Still, it is enough to identify the key drivers of a model prediction.

Indeed, both moth and butterfly communities have been shown to be explorable using our approach. Providing that sufficient inputs (traits and covariates) can be mobilised, good predictive performance is likely observed and explained. For both communities, a dominating effect of wind speed was highlighted which, although not surprising for winged insects, is rarely used as a predictor in prior similar studies (Bell et al., 2020; Ovaskainen

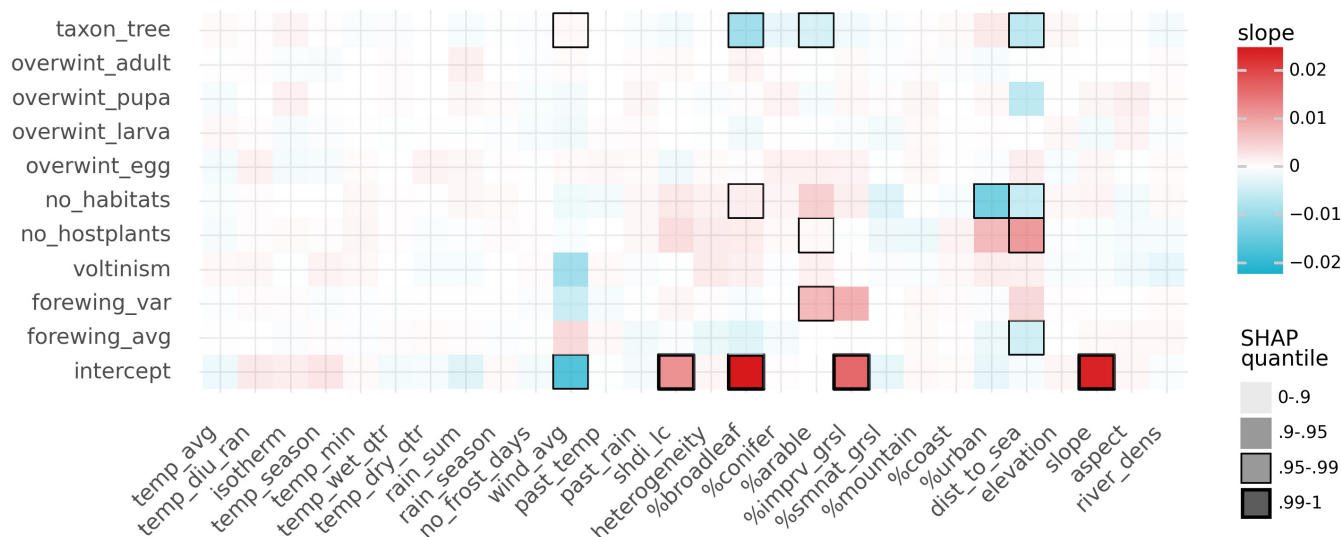




**FIGURE 5** Butterflies' species-specific direct responses to the environmental covariates. The 1%, 5% and 10% most important inputs are highlighted with darker borders and shades. See covariates meanings in Appendix E.

et al., 2016; Palmer et al., 2017; Roy et al., 2001). It is possible that wind only becomes a strong predictor through trait mediation and nonlinearity, in which case, studies building on linear models would necessarily not capture that relationship. The other decisive predictors—namely the proportions of broadleaf woodland, arable land and improved grassland—are common drivers of studies on insect declines (see e.g. Bell et al., 2020; Blumgart et al., 2022; Ovaskainen et al., 2016). Both moths and butterflies appear here significantly affected by those predictors in ways that not necessarily involves trait mediation.

One limit about the present study is that SDMs only concern environmental filtering, that builds on the fundamental niche of a species but is only one cause of the species distributions patterns. Others are dispersal and biotic interactions (e.g. competitive exclusion), which further shape the species distribution by defining their realised niche (Kraft et al., 2015; Poggiato et al., 2021). Dispersal cannot explicitly be accounted for in a network such as ours, with tabular inputs, in which no nonlocal effect is encoded. Yet, using distance buffers (as in Hengl et al., 2018) as inputs in the species-specific branch, we hope



**FIGURE 6** Moths' trait-mediated shared responses to the environmental covariates. The 1%, 5% and 10% most important inputs are highlighted with darker borders and shades. See covariates meanings in Appendix E.

that elements of the spatial structure of a given species distribution, unexplained by the spatial structure of the other environmental features, can be captured as proximity effects. Figure 5 and Figure 11 (Appendix D) show that most species have strong responses to the distance buffer inputs (column *spatial*), suggesting that significant residual spatial structure is identified. Its origin can be either the spatial structure of missing important predictors, or signals of population redistribution processes. In the same way, residual species correlation (or association) can be a sign of biotic interactions (Pollock et al., 2014), and are essential to JSDMs' ability to make joint multivariate predictions (something our ANN does not feature). However, even if facilitative effects among moths or butterflies exist, positive residual correlations more likely suggest that the model misses significant environmental covariates (Poggiato et al., 2021), like the presence of a common predator.

## 5 | CONCLUSIONS

We have demonstrated that a trait-based approach can be encoded as an ANN by using a time distributed layer. This enables machine learning models to be used for identifying community-wide shared response, making them suitable in practice for trait-based MSDMs, when they were previously limited to Stacked SDMs. In addition, because explainability is key to any SDM application, we provide a means to visualise species specific and shared responses to the environment. Our solution builds on the SHAP package to open the black box, hence lifting another obstacle in using machine learning for SDMs. Our illustrative case studies show better performance and tractability than existing methods, as well as highlighting decisive drivers of butterfly and moth community composition. An immediate perspective of the present work is its application to more

insect communities, with the hope of gaining further understanding of biodiversity shifts and allowing more accurate forecasting of the impacts of the drivers of changes.

## AUTHOR CONTRIBUTIONS

Yoann Bourhis conceived the ideas and designed methodology; James Bell and Chris Shortall collected the data; All authors contributed to the data analysis; Yoann Bourhis led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

## ACKNOWLEDGEMENTS

This study is part of NE/V00686X/1 Drivers and Repercussions of UK Insect Declines (DRUID). The Rothamsted Insect Survey, a National Capability, is funded by the Biotechnology and Biological Sciences Research Council (BBSRC) under the Core Capability Grant BBS/E/C/000.J0200. We also thank Tom Oliver, Luke Evans and Chris Hassall for fruitful discussions along the design of this study. This study builds on UK Butterfly Monitoring Scheme (UKBMS) data copyright and database right of Butterfly Conservation, the UK Centre for Ecology & Hydrology, British Trust for Ornithology, and the Joint Nature Conservation Committee, 2020. We are also thankful to the anonymous reviewers for their inputs, especially for the suggestion of merging our two branches using an ensemble.

## CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14097>.



## DATA AVAILABILITY STATEMENT

The model inputs for the Moths (Species occurrences, traits and environmental covariates) are available for [download](#) under GNU GPL4 licence (Bourhis et al., 2022). They are accompanied by the UK grid at 1 km scale of the environmental covariates for the year 2020 to allow for the reproduction the predictions shown in this study. The python code to train and run the ANN is available on [gitlab](#) and distributed under GNU GPL3 licence.

## ORCID

Yoann Bourhis  <https://orcid.org/0000-0002-9365-9597>

James R. Bell  <https://orcid.org/0000-0003-1640-6120>

Chris R. Shortall  <https://orcid.org/0000-0002-7175-5393>

William E. Kunin  <https://orcid.org/0000-0002-9812-2326>

Alice E. Milne  <https://orcid.org/0000-0002-4509-0578>

## REFERENCES

- Baker, R. E., Peña, J.-M., Jayamohan, J., & Jérusalem, A. (2018). Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biology Letters*, 14(5), 20170660.
- Bell, J. R., Blumgart, D., & Shortall, C. R. (2020). Are insects declining and at what rate? An analysis of standardised, systematic catches of aphid and moth abundances across Great Britain. *Insect Conservation and Diversity*, 13(2), 115–126.
- Blumgart, D., Botham, M. S., Menéndez, R., & Bell, J. R. (2022). Moth declines are most severe in broadleaf woodlands despite a net gain in habitat availability. *Insect Conservation and Diversity*, 15(5), 496–509. <https://doi.org/10.1111/icad.12578>
- Bourhis, Y., Bell, J. R., Shortall, C. R., & Milne, A. E. (2022). *Yearly occurrence of 544 species of moths (UK 1990–2019), with trait values and putative environmental drivers*. Rothamsted Research.
- Brereton, T., Botham, M., Middlebrook, I., Randle, Z., Noble, D., Harris, S., Dennis, E., Robinson, A., Peck, K., & Roy, D. (2020). *United Kingdom Butterfly Monitoring Scheme report for 2019*. Technical report, UK Centre for Ecology & Hydrology, Butterfly Conservation, British Trust for Ornithology and Joint Nature Conservation Committee.
- Bystrova, D., Poggiato, G., Bektaş, B., Arbel, J., Clark, J. S., Guglielmi, A., & Thuiller, W. (2021). Clustering species with residual covariance matrix in joint species distribution models. *Frontiers in Ecology and Evolution*, 9, 601384. <https://doi.org/10.3389/fevo.2021.601384>
- Calabrese, J. M., Certain, G., Kraan, C., & Dormann, C. F. (2014). Stacking species distribution models and adjusting bias by linking them to macroecological models. *Global Ecology and Biogeography*, 23(1), 99–112. <https://doi.org/10.1111/geb.12102>
- Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, 21(1), 6.
- Cook, P., Tordoff, G., Davis, A., Parsons, M., Dennis, E., Fox, R., Botham, M., & Bourn, N. (2021). Traits data for the butterflies and macro-moths of Great Britain and Ireland. *Ecology*, 103, e3670.
- Deneu, B., Servajean, M., Bonnet, P., Botella, C., Munoz, F., & Joly, A. (2021). Convolutional neural networks improve species distribution modelling by capturing the spatial structure of the environment. *PLoS Computational Biology*, 17(4), e1008856. <https://doi.org/10.1371/journal.pcbi.1008856>
- Dennis, E. B., Morgan, B. J. T., Freeman, S. N., Brereton, T. M., & Roy, D. B. (2016). A generalized abundance index for seasonal invertebrates. *Biometrics*, 72(4), 1305–1314.
- Dolédec, S., Chessel, D., ter Braak, C. J. F., & Champely, S. (1996). Matching species traits to environmental variables: A new three-table ordination method. *Environmental and Ecological Statistics*, 3(2), 143–166.
- Dray, S., Choler, P., Dolédec, S., Peres-Neto, P. R., Thuiller, W., Pavoine, S., & Braak, C. J. F. T. (2014). Combining the fourth-corner and the RLQ methods for assessing trait responses to environmental variation. *Ecology*, 95(1), 14–21.
- Fernández, A. (2018). *Learning from imbalanced data sets* (1st ed.). Springer.
- Fox, R., Dennis, E. B., Harrower, C., Blumgart, D., Bell, J., Cook, P., Evans-Hill, L., Haynes, F., Hill, D., Isaac, N., Parsons, M., Pocock, M., Prescott, T., Randle, Z., Shortall, C., Tordoff, G., Tudson, D., & Bourn, N. (2020). *The state of Britain's larger moths 2021*. Technical report, Butterfly Conservation, Rothamsted Research and UK Centre for Ecology & Hydrology.
- Guisan, A., & Rahbek, C. (2011). SESAM—A new framework integrating macroecological and species distribution models for predicting spatio-temporal patterns of species assemblages. *Journal of Biogeography*, 38(8), 1433–1444. <https://doi.org/10.1111/j.1365-2699.2011.02550.x>
- Harris, D. J. (2015). Generating realistic assemblages with a joint species distribution model. *Methods in Ecology and Evolution*, 6(4), 465–473. <https://doi.org/10.1111/2041-210X.12332>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, 6, e5518.
- Hill, L., Hector, A., Hemery, G., Smart, S., Tanadini, M., & Brown, N. (2017). Abundance distributions for tree species in Great Britain: A two-stage approach to modeling abundance using species distribution modeling and random forest. *Ecology and Evolution*, 7(4), 1043–1056. <https://doi.org/10.1002/ece3.2661>
- Hollis, D., McCarthy, M., Kendon, M., Legg, T., & Simpson, I. (2019). HadUK-grid—A new UK dataset of gridded climate observations. *Geoscience Data Journal*, 6(2), 151–159. <https://doi.org/10.1002/gdj3.78>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366.
- Kempel, A., Razanajatovo, M., Stein, C., Unsicker, S. B., Auge, H., Weisser, W. W., Fischer, M., & Prati, D. (2015). Herbivore preference drives plant community composition. *Ecology*, 96(11), 2923–2934. <https://doi.org/10.1890/14-2125.1>
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., & Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Functional Ecology*, 29(5), 592–599. <https://doi.org/10.1111/1365-2435.12345>
- Legendre, P., Galzin, R., & Harmelin-Vivien, M. L. (1997). Relating behavior to habitat: Solutions to Thefourth-corner problem. *Ecology*, 78(2), 547–562. <https://doi.org/10.1890/0012-9658%281997%29078%5B0547%3ARBTHTST%5D2.0.CO%3B2>
- Lever, J., Krzywinski, M., & Altman, N. (2016). Classification evaluation. *Nature Methods*, 13(8), 603–604. <https://doi.org/10.1038/nmeth.3945>
- Lundberg, S., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv:1705.07874 [cs, stat]*. arXiv: 1705.07874.
- Middleton-Welling, J., Dapporto, L., García-Barros, E., Wiemers, M., Nowicki, P., Plazio, E., Bonelli, S., Zaccagno, M., Šašić, M., Liparova, J., Schweiger, O., Harpke, A., Musche, M., Settele, J., Schmucki, R., & Shreeve, T. (2020). A new comprehensive trait database of European and Maghreb butterflies, Papilionoidea. *Scientific Data*, 7(1), 351.
- Molnar, C. (2021). *Interpretable machine learning*. Independently Published. <https://www.christophm.github.io/interpretable-ml-book/>
- Morton, R., Marston, C., O'Neil, A., & Rowland, C. (2020). *Land cover map 2019*. Technical report, NERC Environmental Information Data Centre.
- Münkemüller, T., Gallien, L., Pollock, L. J., Barros, C., Carboni, M., Chalmardier, L., Mazel, F., Mokany, K., Roquet, C., Smyčka, J., Talluto,

- M. V., & Thuiller, W. (2020). Dos and don'ts when inferring assembly rules from diversity patterns. *Global Ecology and Biogeography*, 29(7), 1212–1229. <https://doi.org/10.1111/geb.13098>
- Nielsen, M. A. (2015). *Neural networks and deep learning*. Determination Press.
- Ovaskainen, O., & Abrego, N. (2020). *Joint species distribution modelling: With applications in R*. Ecology, Biodiversity and Conservation. Cambridge University Press.
- Ovaskainen, O., Roy, D. B., Fox, R., & Anderson, B. J. (2016). Uncovering hidden spatial structure in species communities with spatially explicit joint species distribution models. *Methods in Ecology and Evolution*, 7(4), 428–436.
- Ovaskainen, O., Tikhonov, G., Norberg, A., Guillaume Blanchet, F., Duan, L., Dunson, D., Roslin, T., & Abrego, N. (2017). How to make more out of community data? A conceptual framework and its implementation as models and software. *Ecology Letters*, 20(5), 561–576.
- Palmer, G., Platts, P. J., Brereton, T., Chapman, J. W., Dytham, C., Fox, R., Pearce-Higgins, J. W., Roy, D. B., Hill, J. K., & Thomas, C. D. (2017). Climate change, climatic variation and extreme biological responses. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 372(1723), 20160144.
- Pichler, M., & Hartig, F. (2021). A new joint species distribution model for faster and more accurate inference of species associations from big community data. *Methods in Ecology and Evolution*, 12(11), 2159–2173. <https://doi.org/10.1111/2041-210X.13687>
- Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the interpretations of joint modeling in community ecology. *Trends in Ecology & Evolution*, 36(5), 391–401.
- Pollock, L. J., Morris, W. K., & Vesk, P. A. (2012). The role of functional traits in species distributions revealed through a hierarchical model. *Ecography*, 35(8), 716–725.
- Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., Vesk, P. A., & McCarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a joint species distribution model (JSDM). *Methods in Ecology and Evolution*, 5(5), 397–406. <https://doi.org/10.1111/2041-210X.12180>
- Roy, D. B., Rothery, P., Moss, D., Pollard, E., & Thomas, J. A. (2001). Butterfly numbers and weather: Predicting historical trends in abundance and the future effects of climate change. *Journal of Animal Ecology*, 70(2), 201–217.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3), e0118432.
- Storkey, J., Macdonald, A. J., Bell, J. R., Clark, I. M., Gregory, A. S., Hawkins, N. J., Hirsch, P. R., Todman, L. C., & Whitmore, A. P. (2016). Chapter one—The unique contribution of Rothamsted to ecological research at large temporal scales. In A. J. Dumbrell, R. L. Kordas, & G. Woodward (Eds.), *Advances in ecological research, volume 55 of Large-scale ecology: Model systems to global perspectives* (pp. 3–42). Academic Press.
- Thomas, J., & Lewington, R. (2019). *The butterflies of Britain & Ireland*. Bloomsbury Wildlife, revised edition.
- Tikhonov, G., Opedal, Ø. H., Abrego, N., Lehikoinen, A., de Jonge, M. M. J., Oksanen, J., & Ovaskainen, O. (2020). Joint species distribution modelling with the r-package Hmsc. *Methods in Ecology and Evolution*, 11(3), 442–447. <https://doi.org/10.1111/2041-210X.13345>
- Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. *Diversity and Distributions*, 15(4), 565–576. <https://doi.org/10.1111/j.1472-4642.2009.00567.x>
- Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2020). Improving prediction of rare species' distribution from community data. *Scientific Reports*, 10(1), 12230.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**Appendix A:** Competing architectures.

**Appendix B:** Traits principal component analysis

**Appendix C:** Moth predictions.

**Appendix D:** Moth species-specific response.

**Appendix E:** Environmental covariates.

**How to cite this article:** Bourhis, Y., Bell, J. R., Shortall, C. R., Kunin, W. E., & Milne, A. E. (2023). Explainable neural networks for trait-based multispecies distribution modelling—A case study with butterflies and moths. *Methods in Ecology and Evolution*, 14, 1531–1542. <https://doi.org/10.1111/2041-210X.14097>