RESEARCH ARTICLE

Extending vegetation site data and ensemble models to predict patterns of foliage cover and species richness for plant functional groups

Megan J. McNellie · Ian Oliver · Simon Ferrier · Graeme Newell · Glenn Manion · Peter Griffioen · Matt White · Terry Koen · Michael Somerville · Philip Gibbons

Received: 4 July 2019 / Accepted: 15 February 2021 / Published online: 6 March 2021
© Crown 2021

Abstract

Context Ensembles of artificial neural network models can be trained to predict the continuous characteristics of vegetation such as the foliage cover and species richness of different plant functional groups. **Objectives** Our first objective was to synthesise existing site-based observations of native plant species to quantify summed percentage foliage cover and species richness within four functional groups and in totality. Secondly, we generated spatially-explicit,

continuous, landscape-scale models of these functional groups, accompanied by maps of the model residuals to show uncertainty.

Methods Using a case study from New South Wales, Australia, we aggregated floristic observations from 6806 sites into four common plant growth forms (trees, shrubs, grasses and forbs) representing four different functional groups. We coupled these response data with spatially-complete surfaces describing environmental predictors and predictors that reflect landscape-scale disturbance. We predicted the distribution of foliage cover and species richness of these four plant functional groups over 1.5 million

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10980-021-01221-x>.

M. J. McNellie (✉)
Department of Planning, Industry and Environment,
PO Box 5336, Wagga Wagga, NSW 2650, Australia
e-mail: megan.mcnellie@environment.nsw.gov.au

M. J. McNellie · P. Gibbons
Fenner School of Environment and Society, The
Australian National University, Frank Fenner Building,
Linnaeus Way, Acton, ACT 2601, Australia
e-mail: philip.gibbons@anu.edu.au

I. Oliver
Department of Planning, Industry and Environment,
Gosford, NSW 2250, Australia
e-mail: ian.oliver2@environment.nsw.gov.au

S. Ferrier
CSIRO Land and Water, PO Box 1700, Canberra,
ACT 2601, Australia
e-mail: simon.ferrier@csiro.au

G. Newell · P. Griffioen · M. White
Department of Environment, Land, Water and Planning,
Arthur Rylah Institute for Environmental Research, 123
Brown Street, Heidelberg, VIC 3084, Australia
e-mail: graeme.newell@delwp.vic.gov.au

P. Griffioen
e-mail: peter.griffioen@delwp.vic.gov.au

M. White
e-mail: matt.white@delwp.vic.gov.au

G. Manion · M. Somerville
Department of Planning, Industry and Environment,
Armidale, NSW 2351, Australia
e-mail: manionglenn@bigpond.com

M. Somerville
e-mail: msomerv2@gmail.com

hectares. Importantly, we display spatially explicit model residuals so that end-users have a tangible and transparent means of assessing model uncertainty.

Results Models of richness generally performed well (R^2 0.43–0.63), whereas models of cover were more variable (R^2 0.12–0.69). RMSD ranged from 1.42 (tree richness) to 29.86 (total native cover). MAE ranged from 1.0 (tree richness) to 20.73 (total native foliage cover).

Conclusions Continuous maps of vegetation attributes can add considerable value to existing maps and models of discrete vegetation classes and provide ecologically informative data to support better decisions across multiple spatial scales.

Keywords Growth form · Residual error · Ensemble · Neural network · Predictive modelling · Site-based floristic records · Spatially-explicit vegetation models · Vegetation richness · Vegetation cover

Introduction

Landscape ecology integrates ecological patterns and processes across multiple species and spatial scales (Turner 2005). Integrating ecologically important attributes across landscape scales can enhance conservation and land management decision making (Margules and Pressey 2000; Pressey et al. 2007; Ferrier and Drielsma 2010). Often maps of vegetation extent and community type are the primary sources of spatial information used to represent patterns of ecologically important attributes and underpin decision-making across multiple spatial scales. Typically, maps of vegetation extent depict binary classifications such as ‘extant versus cleared’, ‘woody versus non-woody’ or ‘native versus non-native’. Maps of vegetation communities often depict discrete boundaries between community types, each represented as internally homogeneous composition, structure or function (Noss 1990).

However, discrete boundaries seldom represent continuous characteristics or the functional status of vegetation, such as the presence, abundance or diversity of different functional groups (Noss 1990; Evans and Cushman 2009; De Cáceres and Wiser 2011). To overcome some of these limitations, maps that represent the continuous variation and gradients in important vegetation attributes have been increasingly being created and used in landscape ecology (Austin and Smith 1989; DeFries et al. 1995; Pausas and Austin 2001). Continuous vegetation models recognise that patterns in vegetation often present as gradients and typically rely on remote sensing. Remote sensing data have advanced the development and production of continuous vegetation models by integrating biochemical, physiological and structural quantities of vegetation across a range of spatial and temporal scales (Houborg et al. 2015; Lausch et al. 2020). However, due to confounding and complex interactions between leaf, canopy, atmosphere and reflectance, integrating remote sensing data in ecology remains challenging with large uncertainties (Schimel et al. 2020; Schrodtt et al. 2020), especially when used for informing models that target species or aggregated functional groups (Ustin and Gamon 2010).

In contrast to approaches that rely on remote sensing to map important vegetation attributes, we have focussed on constructing spatially explicit maps of two important continuous characteristics of vegetation—foliage cover and species richness—from existing on-ground species observations. Modelling continuous surfaces from these site-based observations has, to date, focused mostly on individual species distribution models (i.e. predict first, assemble later strategy Ferrier and Guisan 2006) or on modelling overall native species richness (i.e. assemble first, predict later strategy Ferrier and Guisan 2006). Our approach extends the assemble first, predict later strategy to work not just with overall native cover or richness, but with the foliage cover and species richness of functional groups, described by growth forms, thereby providing more information on important continuous characteristics of vegetation.

Aggregations of plant species based on their growth form represent a long-standing (Warming and Vahl 1909) and simple (Kattge et al. 2020) approach for characterising plant functional groups and thereby helping to reveal spatial patterns in habitats, resources and communities at a range of spatial scales.

T. Koen
Department of Planning, Industry and Environment,
Cowra, NSW 2794, Australia
e-mail: sandtkoen@bigpond.com

Functional groups are useful for consolidating and describing the multi-layer structural complexity within vegetation communities. These aggregations also provide a more nuanced interpretation of the relative abundance and composition of different groups of plant species within communities (Warming and Vahl 1909; Box 1981), the effects of disturbance and competition (Grime 1977), dispersal and migration (Svenning and Sandel 2013) and adaptations to climate (Smith 1913; Cain 1950). Globally, site-based floristic inventories that record individually observed plant species (Franklin et al. 2017; Bruelheide et al. 2019) and their growth form allocations (Engemann et al. 2016; Oliver et al. 2019; Kattge et al. 2020) are becoming easier to access. These accessible databases offer opportunities to explore the emergent properties of plant functional groups to inform the ecological patterns and processes that underpin landscape ecology across multiple species and spatial scales.

Here, we focus on tree, shrub, grass and forb growth forms to demonstrate an alternative approach to modelling spatially-explicit and continuous characteristics of foliage cover and species richness of different plant functional groups. The first objective of our study was to aggregate site-based observations of native species into the four growth form based functional groups. We then calculated the foliage cover and species richness for each functional group. We coupled these site-based observations with spatially-complete surfaces (hereafter referred to as predictor surfaces, also defined as environmental covariates or explanatory variables) that influence the growth and morphology of vegetation, as well as landscape-scale disturbance. The second objective was to extrapolate the spatial patterns in the foliage cover and species richness of plant functional groups using ensembles of artificial neural networks (ANN) models and produce maps of these predictions across the landscape. Our study demonstrates a rigorous and repeatable approach to producing spatially-explicit and continuous maps of foliage cover and species richness of different plant functional groups for conservation and land management applications.

Methods

Study area

Our study area is located in north-eastern New South Wales (NSW), Australia (Fig. 1). Vegetation is diverse, ranging from Closed Rainforests in the eastern elevated (1510 m) regions to Arid Shrublands and Grasslands on the drier western plains (Keith 2004). The area is dominated by privately-owned land used for agriculture (52% land used for grazing and 23% used for cultivation, including irrigated cropping). Less than 10% of the area is protected within National Park or public reserve.

Site-based data to inform landscape-scale models: the response variables

We extracted floristic records from in the BioNet Database (<http://www.bionet.nsw.gov.au>). Only records that were: observed in a fixed area (0.04 ha); included either a percentage foliage cover estimate (< 1% to 100%) or a cover-abundance score (e.g. 1–6 Braun-Blanquet cover-abundance score) for every plant species and contained all other required metadata information (Step 1.1 in Fig. 2) were used as response data. The suite of sites extracted from the database spanned a survey period from 1986 to 2011. Site locations were also cross-checked with a map of extant vegetation (NSW Office of Environment and Heritage 2017) to assess if native vegetation had been removed or cleared subsequent to the floristic survey. Because our predictor surfaces were resampled to 25 m resolution (see *Environmental and disturbance variables* below), we ensured that sites had a minimum of 55 m to the nearest neighbour (McNellie et al. 2015). Where sites were located within 55 m of each other, one of the sites was randomly selected and excluded from further analyses. Floristic data recorded on an ordinal cover-abundance scale were converted to percentage cover values following methods described in McNellie et al. (2019) (Step 1.2 in Fig. 2). All species were assigned as either native or non-native (Harden 1990). We accessed an existing framework for allocating native species to growth form (Oliver et al. 2019) (Step 1.3 in Fig. 2). We focused on the four dominant growth form groups: trees, shrubs, grass and grass-like (hereafter referred to as grasses) and forbs. Nativeness and growth form were used to define functional

Fig. 1 Location of the study area (approximately 115 000km²) showing topographic relief—northern New South Wales, Australia

groups. For each functional group, cover was calculated by summing the foliage cover estimates of all native species and richness was assessed by counting all native species (after aggregating subspecies and varieties). In total, we synthesised ten sets of response data: total native cover; total native richness; as well as foliage cover and species richness of trees, shrubs, grasses and forbs.

To improve predictive accuracy and to build a representative dataset, we created an additional set of background points (Step 1.4 in Fig. 2). Given our exclusive interest in the foliage cover and species richness of the plant functional groups, our background points were locations where there was virtually no perennial cover of native vegetation. Background points were important in this study as the training data were almost exclusively obtained from natural or near-natural vegetation and the objective of this study was to extrapolate our models across all land tenures. This complementary set of background points were

generated using on-screen digitising point registration using satellite imagery (2009 SPOT 5) as a backdrop to identify under-sampled land uses that were anthropogenically modified and contained no native terrestrial vegetation such as irrigation bays, water bodies and infrastructure. All background points were at least 100 m apart to ensure they were not near-neighbours. Each background point was attributed as having 0% foliage cover and zero native species richness. The dataset used to train the models contained 6806 sites and an additional 2462 background points. The subset of sites that matched the temporal period of the remote sensing variables (2005 to 2012 and that were used to predict cover—detailed below) was 3021 floristic sites and we selected a random subset of 1015 background points. The training matrix, absence sites and extracted values for all predictor surfaces are stored at <https://doi.org/10.6084/m9.figshare.13728301>.

STEP 1. Assimilate site-scaled data to inform landscape scaled models – the response variables **STEP 2. Assimilate environmental and disturbance gradients – the predictor surfaces** **STEP 3. Train artificial neural network models** **STEP 4. Predict across the landscape** **STEP 5. Represent model predictions and residual error as raster maps**

◀ **Fig. 2** Outline the five components of this study: (i) transforming site-based data into vegetation attributes; (ii) sourcing spatial layers that represent environmental and disturbance gradients; (iii) training artificial neural networks (ANN) models; (iv) using the trained ANN model to predict attributes across the whole landscape; and (v) rendering the average results from ensembles of predicted models into a spatially explicit map for each attribute, accompanied by the standardised residual error

Environmental and disturbance variables: the predictor surfaces

We selected a suite of predictor surfaces that describe spatial variation in (a) environmental attributes that directly (temperature, moisture, radiation, soil) or indirectly (geology, topography) influence the resources and conditions controlling growth and morphology of vegetation (Box 1981; Guisan and Zimmermann 2000; Pressey et al. 2000; Austin 2002; Franklin 2009) and (b) disturbance variables that modify and fragment vegetation, based on an a priori assessment of the potential importance of these variables (Step 2.1 in Fig. 2 and Supplementary Material S1).

Abiotic environmental surfaces

Climatic and topographic surfaces were developed using the Australian 1 s, smoothed digital elevation model (DEM-S) (Gallant et al. 2011). Raster surfaces were resampled to 25 m resolution to match the observational scale of the response data (Williams et al. 2012). Climatic variables were calculated using the MTHCLIM module in ANUCLIM v6.1 (Xu and Hutchinson 2013) for the 1921–1995 epoch. Detailed and additional information on predictor surfaces is provided in Supplementary Material S1.

Contemporary disturbance surfaces

Anthropogenic modification has a direct link with changes and loss in biodiversity (Maxwell et al. 2016), and land-use mapping at the catchment-scale can be used as a proxy for disturbance (e.g. Gardner et al. 2009). Catchment-scale mapping describes the different primary management practices across the broader landscape and was captured using a consistent framework at a continental-scale as per standards set by the Australian Collaborative Land Use Mapping Program

(ACLUMP) and using the Australian Land Use and Management ALUM Classification (Lesslie et al. 2006). We used the ALUM framework to assign mapped land use classes to seven major groups: conservation areas that are essentially natural ecosystems; tree and shrub cover on private land; grazing; cropping and horticulture; land uses subject to extensive modification (which included all urban areas, roads, mining, power generation and areas used for intensive animal production); rivers and wetlands (see Supplementary Material S1).

Predictor surfaces derived from Landsat TM imagery were used only to inform models of total native vegetation cover and vegetation cover by growth form. Landsat TM imagery from 2005 to 2012 was used to calculate normalised difference vegetation index (NDVI), foliage projected cover (FPC) (Lucas et al. 2006) and bare ground including exposed soil and rock (Scarth et al. 2010). Consequently, the response data used to train and verify models of total native vegetation cover and vegetation cover by growth form were a subset of sites surveyed between 2005 and 2012. Prior to training the models, we tested for collinearity between the predictors by calculating the Variance Inflation Factor (VIF). Where the VIF exceeded 5, the predictor was excluded and the VIF was recalculated (O'Brien 2007) (Supplementary Material S2 details which variables were excluded).

Modelling framework

Of the range of modelling methods available to extrapolate site-based observation data (Elith et al. 2006), we used ensembles of ANN models (see additional information on the model parameters detailed in Supplementary Material S3). ANNs are advantageous for ecological applications where data do not meet parametric statistical assumptions and the relationships between the response data and the predictor surfaces are complex, unknown or non-linear (Bishop 1995; Fielding 1999). We chose ANN (Statistica software v10) because they are effective for resolving complex predictions and can handle redundant or co-linear predictor variables (Statsoft Inc. 2013). In addition, other non-parametric models, such as Random Forests, interpolate by recursively partitioning the data, whereas ANNs can extrapolate which

is more useful in predicting lower and higher ends of the data distribution (Heikkinen et al. 2012).

For each vegetation attribute, the training matrix (Step 2.2 in Fig. 2) was randomly split into three subsets: 50% of the data were used to train the model; 20% were used to test the model; the remaining 30% were withheld from training and testing and were used as independent data to verify the model (Step 3.1 in Fig. 2). This form of three-way data partitioning is a rigorous method for validating models (Fielding 1999; Chicco 2017; Quinn et al. 2021). We trained 25 ANN models and averaged them to produce an ‘ensemble of random samples’ model (Opitz and Maclin 1999; Araújo and New 2007; Shmueli 2010). Before training neural networks, input and target variables were scaled using linear transformations such that the original minimum and maximum of every variable were mapped to the range (0, 1). The predictive performance of each ensemble model was evaluated by calculating the coefficient of determination (R^2) (Step 3.2 in Fig. 2). It is important to note that model performance was judged by determining how well the model performed when applied to new data (the out-of-sample subset). Parity between the R^2 for the training and hold-out subsets indicates how well the random sub-sampling has represented the range of variation in the entire dataset (Step 3.3 in Fig. 2).

We compared the root mean squared deviation (RMSD) (see Supplementary S5) as an estimate of the deviation of the transformed cover values from the 1:1 line. We also calculated the mean absolute error (MAE) (see Supplementary S5). Both error estimates report errors in the same scale as the input data.

Sensitivity analyses were performed to assess the contribution of each predictor surface to the model (Step 3.3 in Fig. 2 and Supplementary Material S6). Sensitivity analyses are unit-less measurements and show how the model performs when each predictor is removed from the analysis.

Spatial autocorrelation arises in ecological data because points that are nearer to one another are more likely to be similar, in either the response variable (e.g. occurrence or abundance) or their environmental profile of the predictor surfaces used in the model, than points that are further apart (Keitt et al. 2002; Dormann et al. 2007; Getis and Ord 2010). To check for the presence of spatial autocorrelation, we assessed the spatial relationship of model residuals (see Step 5.2 in Fig. 2 and Supplementary Material S7) by

evaluating Moran’s Index, z-scores and p-values using ArcGIS Spatial Statistics toolbox (ArcGIS v10.4). Both z-scores and p-values are associated with the standard normal distribution where z-scores are standard deviations (see Supplementary Material S7).

Predicting spatial patterns across the whole landscape

In this stage, we used raster analyses to treat every 100 m grid cell in the study area (approximately 11.5 million grid cells) as a new, unknown, or unsurveyed site (Step 4.1 in Fig. 2). At the centroid of each grid cell (see Supplementary Material S4—NN Extractor.exe) and the underlying values of the predictor surfaces were extracted to build the prediction matrix (see Supplementary Material S4—NN Extract from Binaries.exe for details of the software designed to build this matrix). This matrix (sites $\times$ predictor variables) formed the ‘predict later’ (Ferrier and Guisan 2006) input table used to predict the trends and patterns learned from the training data site (Step 4.2 in Fig. 2).

To build predictive models of each of the ten vegetation attributes, each of the 25 training models per attribute (representing 25 trained networks) were deployed to every grid cell in the prediction matrix which represented the new, unknown cases. These analyses produced 25 prediction models. The final predicted output for each grid cell was averaged to create a single ensemble model for each vegetation attribute (Step 4.3 in Fig. 2). Custom software (Step 5.1 in Fig. 2) (see Supplementary Material S4—NN Tools Prediction.exe) transformed the model output as raster maps (Step 5.2 in Fig. 2). An overlay of mapped non-native vegetation was used to mask out areas that have been identified as cleared of native vegetation (NSW Office of Environment and Heritage 2017).

Representing the model residuals as maps

To investigate the spatial patterns in the model residuals, we used ordinary kriging (implemented in the Spatial Analysis toolbox in ArcGIS v10.4). Ordinary kriging interpolates the unknown locations by considering both the distance and the degree of variation between known data points. This technique presumes each known data point to be locally influenced and this influence decreases as the distance from

the sampled location increases (Paramasivam and Venkatramanan 2019). The interpolation procedure assigns greater weights to closer points. The standardised model residuals were used as known input points and the 12 nearest points were selected using a variable search radius. The interpolation methods used the weighted average to estimate the residual error in unknown locations (see Step 5.3 in Fig. 2). Negative values indicate over-prediction and positive values indicate under-prediction. The standardised residual error can be interpreted like z-scores (standard deviations), whereby values that exceed -2.58 and 2.58 are regarded as spatial outliers. This is a robust method of geostatistical interpolation (Franklin 2009).

Results

Model assessment

For the hold-out subsets, R^2 ranged from 0.69 (total native cover) to 0.12 (for forb cover) (Table 1). When the trained models were tested against the hold-out subsets, the difference in R^2 was small. The difference between the R^2 for the train:test:hold-out subsets for richness models did not exceed 0.05 (tree richness). The R^2 for the train:test:hold-out subsets for cover models did not exceed 0.04 (tree and grass cover). This parity shows that 50:20:30 (train:test:hold-out) partitioning represented the data adequately and that model

performance was equal when applied to the out-of-sample subset.

Both RMSD and MAE are reported in the same units as the response data. The results of the RMSD for the cover models show that the mean deviation of predicted cover relative to the observed cover was small and ranged from 9.1 (forb cover) to 28.95 (total native cover). RMSD values for the richness models show that the predicted richness values relative to observed richness values were also small and ranged from 1.45 (tree richness) to 10.65 (total native richness) (Table 1). MAE for cover models ranged from 5.37 (forb cover) to 20.73 (total native cover). Likewise, MAE for richness models was small and ranged from 1.0 (tree richness) to 7.77 (total native richness) (Table 1). When considered in the context of the potential range of observer error (Cook et al. 2010; Morrison 2016), our models are ecologically realistic and fit for the purpose of assessing foliage cover and species richness of plant functional groups.

The Moran's Index (Supplementary Material S7) suggests that the model residuals were not spatially autocorrelated. Z-scores for the residuals of the cover models ranged between 0.63 (trees) and 0.01 (shrubs and grasses) and the z-scores for the richness models ranged between -0.35 (forbs) and 0.15 (shrubs).

Table 1 Coefficient of determination (R^2) for the ensemble of random samples for each of the ten vegetation attributes

Vegetation attribute	R^2_{Train}	R^2_{Test}	$R^2_{\text{Hold-out}}$	RMSD	MAE
Tree cover	0.63	0.59	0.59	16.23	10.68
Shrub cover	0.31	0.33	0.31	17.34	11.44
Grass cover	0.31	0.28	0.27	21.53	15.07
Forb cover	0.15	0.13	0.12	9.06	5.37
Total cover	0.72	0.69	0.69	29.86	20.73
Tree richness	0.49	0.53	0.54	1.42	1.00
Shrub richness	0.40	0.44	0.44	3.59	2.53
Grass richness	0.40	0.43	0.44	4.00	2.95
Forb richness	0.40	0.42	0.43	5.58	4.06
Total richness	0.59	0.62	0.63	10.68	7.77

Numbers in bold highlight results for the hold-out subset

The root mean squared difference (RMSD) and mean absolute error (MAE) estimates show the mean deviation of predicted cover with respect to the observed cover and predicted richness values with respect to the observed richness values. Number of observations cover models $n = 3021$ and total number of observations for richness models $n = 9268$

Predictor variables and sensitivity analysis

Sensitivity analysis showed that the categorical predictors (land use and great soil group) made the two highest contributions to training models (Supplementary Material S6). Overall, the suite of richness models yielded higher sensitivity values for land use and soils, followed by climatic variables (isothermality and precipitation/evaporation) and soil properties (% clay). FPC had higher sensitivity values for the suite of cover models, especially total native cover and tree cover, which is not surprising given it is a remotely sensed predictor. However, for some of the cover attributes, such as forb cover, there were few strong predictor variables and the resultant predictive model was poor.

Extrapolating spatial patterns across the landscape

We created spatially explicit model-based predictions for ten vegetation attributes. Here we show an example (Fig. 3) of the detailed (1:1 000 000 scale) prediction surface for summed grass cover (%) (Fig. 3a) and the modelled estimate of standardised residual error in the prediction (Fig. 3b), for a section of the study area (Fig. 3c). Standardised residual errors illustrate under- and over-estimation across the entire landscape and are specific to each vegetation attribute. Displaying error maps in this simple form is intended to allow end-users to assess the relative strength of our models at specific locations. To further demonstrate how individual vegetation attributes can be modelled, Fig. 4 shows a detailed example for shrub richness (1:1 000 000 scale) over the same area. The spatially explicit maps for all ten vegetation attributes and their

Fig. 3 Fine-scale representation of predicted grass cover showing the detail of the **a** continuous surface for grass cover (summed %foliage cover), **b** standardised residual error for grass cover and **c** a location diagram within the study area

Fig. 4 Fine-scale representation of predicted shrub richness showing the detail of the **a** continuous surface for shrub richness, **b** standardised residual error for shrub richness and **c** a location diagram within the study area

accompanying residual error maps are supplied as Supplementary Material S8.

Discussion

Extending site-data for predictive ecological modelling

Here we generate continuous predictive vegetation models of foliage cover and species richness for different functional groups. We have shown how existing inventories of floristic data can be assembled into growth forms which can, in turn, be used to map different facets of foliage cover and species richness of plant functional groups at landscape- or regional-scales. The methods we have outlined here have global applications. The ever-growing volume of floristic site

data (Dengler et al. 2011; Schaminée et al. 2011; Peet et al. 2013; Bruelheide et al. 2019) and large-scale syntheses of species to a growth form (Engemann et al. 2016; Oliver et al. 2019; Kattge et al. 2020) has offered opportunities to explore how site-based observations can be extended to inform spatial models. Currently, much effort and attention has been applied to using these floristic data to inventory, describe and map vegetation community types at broad scales (e.g. Mucina and van der Maarel 1989; Grossman et al. 1998; Chytrý et al. 2011; Wiser et al. 2011). Wintle et al. (2005) rank the types of data that are used for modelling habitat suitability. These authors describe data representing counts of individual species as the most difficult to acquire because they are often constrained by cost and time. They also require considerable expertise to identify species. Here we have extracted a large set of existing floristic

observations and used these records to provide an extended, comprehensive and important set of vegetation functional group characteristics that can be used to inform a range of ecological applications.

Our approach of using site-scaled species data collected from a fixed-area (0.04 ha) plots to inform landscape models is useful because species aggregated to growth forms provide very tailored, habitat-specific information for species presence, abundance and distribution. Furthermore, assessments of the spatial patterns of variation in the foliage cover and species richness of functional groups can be critical as they can reveal information about landscape condition and fragmentation (Saunders et al. 1991; Taylor and Lindenmayer 2020) as well as understanding patterns important for predicting the future ecological integrity of native vegetation (Oliver et al. 2021). These models are useful in supporting conservation and land management decision making for multiple species across spatial scales.

Our spatially explicit representation of foliage cover and species richness of individual plant growth forms can overcome some of the sources of variation found in maps of vegetation communities. Hearn et al. (2011) found that most errors in mapping vegetation boundaries were observed where neighbouring vegetation types had similar structure and richness. Delineating communities on a map is often an arbitrary decision that requires a degree of expert interpretation because most vegetation types intergrade across ecotones. Furthermore, Hearn et al. (2011) also found that experts differed in their opinions about which vegetation types were contained within the mapped boundaries, especially when vegetation structure and composition were similar (such as in shrub-dominated heaths or grasslands). Our alternative approach is unimpeded by bounded vegetation categories because we represent foliage cover and species richness as continuous surfaces that can express the heterogeneity in vegetation across the landscape.

Evaluating the predictor surfaces

Land use, as a proxy for disturbance, was a strong predictor for all vegetation cover and richness measures. Land use is a key driver underpinning the modification of natural landscapes (Fischer and Lindenmayer 2007; Newbold et al. 2016). This study addresses the challenge of selecting an appropriate and

comprehensive set of predictor surfaces (Williams et al. 2012) by using predictors that are known to influence growth and morphology of vegetation, such as temperature, light, topography, hydrology (Box 1981; Austin 1998), as well as predictors that are known to modify and fragment landscapes. The approach described here has the capacity to be broadened. Globally, compilations of gridded environmental surfaces that represent climate, soils or topography are readily available (e.g. for climatic data see WorldClim (Fick and Hijmans 2017); for soil data see SoilGrids (Hengl et al. 2017) or digital elevation models from which topographic predictors can be derived (e.g. ASTER Global Digital Elevation Model <https://lpdaac.usgs.gov>). Remote sensing technologies have and will continue to advance vegetation and landscape ecology (Cavender-Bares et al. 2020).

Despite our careful selection of predictor surfaces, some variables will have limitations. Firstly, categorical land use mapping may not represent differing intensity, frequency or duration of disturbance on native vegetation. For example grazing by either native animals or livestock varies in its intensity and impact on native vegetation (Olf and Ritchie 1998; Lunt et al. 2007; Speed and Austrheim 2017; Pausas and Bond 2019) therefore some types of disturbance are not uniform within a single land use category. Secondly, some disturbance events, such as fires, droughts or floods, are stochastic in time (duration, intensity and frequency) and space (scale and extent) (Levin 1992; Lake 2000) and their effect on vegetation can be complex (Pausas and Austin 2001). A static surrogate such as land use may not capture the spatial and temporal variation in disturbance (Drielsma and Ferrier 2006); however, agricultural transformation is a significant driver of biodiversity loss (Maxwell et al. 2016). By focussing on different growth forms, we implicitly acknowledge that different disturbances influence different functional groups in different ways. Advances in data availability (such as routinely updated land use information, or remotely sensed data, including hyperspectral, LiDAR, IKONOS, Quickbird, Landsat ETM and Sentinel-2) may offer opportunities to dynamically and iteratively update and refine categorical variables such as land use (Leitão and Santos 2019), to better predict characteristics of vegetation at a regional scale (Lausch et al. 2020). At a global scale, predictors that relate to land use are often inferred from remotely sensed land cover

data (see Socioeconomic Data and Applications Center <http://sedac.ciesin.columbia.edu>). We also envisage the potential to explore more ecologically relevant predictor variables that may be more suited to explain patterns in forb and grass richness and cover as these smaller growth forms may need to be modelled at a finer scale (Kelemen et al. 2013; Johnson et al. 2020).

Simple and transparent maps of model residuals

One of the substantial benefits of our modelling approach is the mapping of standardised residual error as an expression of model uncertainty. When displayed as a map, the ‘known error’ at every site is used to predict the ‘unknown error’ across the whole landscape. Here we have used ordinary kriging to spatially interpolate the residual error across the whole study area. Kriging is used to interpolate point information (Miller 2005; Sajid et al. 2013), and here we present the residual errors so that end-users can assess where the models have under- or over-predicted in places where they have specific interests, such as inside National Parks or within the expected range of species’ habitats. This is especially useful where end-users may combine the outputs from several models. The spatial predictions for standardised residual error show that some vegetation attributes are likely to be temporally dynamic, such as forb cover and richness, which have a greater range in the residual error. Whereas vegetation attributes that are relatively stable through time, such as tree richness, show a narrower range in the residual error. Predictive models are prone to uncertainty. The sources of model uncertainty can arise from any (or all) of the steps outlined in Fig. 2, including errors and temporal variation in the response data; inaccurate, imprecise or absent predictor surfaces; or the model may fail to adequately associate (learn) the patterns and relationships between the response data and the predictor variables.

Practical applications for mapped vegetation attributes

Quantifying and mapping attributes of plant functional groups, be that of discrete growth forms or in totality, offers an approach for assessing vegetation cover and richness across landscapes or regions. In doing so, we

have extended the assemble-first/predict-later strategy (Ferrier and Guisan 2006) to work not just with overall richness, but with foliage cover and species richness within functional groups which provides more information on the continuous patterns of vegetation form and function. For instance, predictions from these models could be used to inform the habitat preferences for species (e.g. Kissling et al. 2018; Lindenmayer et al. 2018) especially where higher cover or richness of different growth forms contributes to greater habitat complexity (Brown et al. 1995; Rowe and Speck 2005; Ashcroft et al. 2017). These types of biologically-orientated models that can be used to inform habitat-specific occupancy models (McElhinny et al. 2006), especially the foliage cover and species richness of grasses and forbs, which are often overlooked (McElhinny et al. 2005; Gilliam 2007). The attributes of vegetation at a site, such as native species richness underlie conservation strategies (Fleishman et al. 2006) and can be used to identify threats to biodiversity (Hooper et al. 2005; Evans et al. 2011).

In vegetation types where some growth form components are missing, yet are expected, active restoration can be targeted towards extending or repairing these missing components (Banks-Leite et al. 2020; Oliver et al. 2021). We foresee an improved approach to identifying growth forms that could be actively restored to improve the conservation value and enhance some vegetation types. For example, Lindenmayer et al. (2018) found that underplanting woodland vegetation with shrub understorey improves habitat for native birds.

Conclusions

Many models used to inform landscape ecology rely on maps of discrete vegetation entities, yet continuous modelled surfaces can convey considerably more information. Widely available floristic datasets that have recorded individual plant species and their location offer a largely untapped source of information to inform continuous models. To date, modelling and mapping the continuous variation in vegetation has focused mostly on presence/absence data, or individual species distribution models, or on modelling overall species richness. In contrast, our approach considers both foliage cover and species richness within functional groups as separate models, thereby

providing richer information on continuous patterns of vegetation.

Acknowledgements This work was undertaken through collaborative partnerships between former NSW Office of Environment and Heritage and the former Border Rivers – Gwydir and Namoi Catchment Management Authorities. Jillian Thonell, Geoff Horn, Simon Smith and Sarah Hill assisted with data preparation. Clive Hilliker advised on the design of Figure 2. Statsoft User Support Forum provided valuable support and information on STATISTICA and neural network analyses. Wade Blanchet and David Nipperess assisted with methods to accommodate spatial autocorrelation. Samantha Travers assisted with VIF analyses. We are grateful to Martin Dillon, David McNellie, Vivian Silvey and Ashley Sparrow and anonymous referees for all their valuable comments on the manuscript.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Author contributions Conceived the ideas and designed methodology—MM, IO, SF, GN, PGrif, MW, PGib. Extracted response data—MS, MM. Extracted predictor data—MM. Designed software tools—GM. Analysed and interpreted the data—MM, PGrif, TK. Prepared all figures, tables, maps, open-access data deposit—MM. Led the writing of the manuscript—MM. Contributed critically to drafting and revising the manuscript—MM, IO, MW, PGrif, PGib, SF, GN. Final approval for publication—ALL.

References

- Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends Ecol Evol* 22(1):42–47
- Ashcroft MB, King DH, Raymond B, Turnbull JD, Wasley J, Robinson SA (2017) Moving beyond presence and absence when examining changes in species distributions. *Glob Change Biol* 23(8):2929–2940
- Austin MP (1998) An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests. *Ann Mo Bot Gard* 85(1):2–17
- Austin MP (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecol Model* 157(2–3):101–118
- Austin M, Smith T (1989) A new model for the continuum concept. *Vegetatio* 83(1/2):35–47
- Banks-Leite C, Ewers RM, Folkard-Tapp H, Fraser A (2020) Countering the effects of habitat loss, fragmentation, and degradation through habitat restoration. *One Earth* 3(6):672–676
- Bishop CM (1995) Neural networks for pattern recognition. Oxford University Press
- Box EO (1981) Predicting physiognomic vegetation types with climate variables. *Vegetatio* 45(2):127–139
- Brown JH, Mehlman DW, Stevens GC (1995) Spatial variation in abundance. *Ecology* 76(7):2028–2043
- Bruehlheide H, Dengler J, Jiménez-Alfaro B, Purschke O, Hennekens SM, Chytrý M, Pillar VD, Jansen F, Kattge J, Sandel B, Aubin I, Biurrun I, Field R, Haider S, Jandt U, Lenoir J, Peet RK, Peyre G, Sabatini FM, Schmidt M, Schrodt F, Winter M, Acic S, Agrillo E, Alvarez M, Ambarlı D, Angelini P, Apostolova I, Arfin Khan MAS, Arnst E, Attorre F, Baraloto C, Beckmann M, Berg C, Bergeron Y, Bergmeier E, Bjorkman AD, Bondareva V, Borchardt P, Botta-Dukát Z, Boyle B, Breen A, Brisse H, Byun C, Cabido MR, Casella L, Cayuela L, Černý T, Chepinoga V, Csiky J, Curran M, Čuštěrevska R, Stevanović ZD, De Bie E, De Ruffray P, De Sanctis M, Dimopoulos P, Dressler S, Ejrnæs R, El-Sheikh MAERM, Enquist B, Ewald J, Fagúndez J, Finckh M, Font X, Forey E, Fotiadis G, García-Mijangos I, Gasper AL, Golub V, Gutierrez AG, Hatim MZ, He T, Higuchi P, Holubová D, Hölzel N, Homeier J, Indreica A, Gürsoy DI, Jansen S, Janssen J, Jedrzejek B, Jiroušek M, Jürgens N, Kącki Z, Kavgacı A, Kearsley E, Kessler M, Knollová I, Kolomiychuk V, Korolyuk A, Kozhevnikova M, Kozub Ł, Krstonošić D, Kühl H, Kühn I, Kuzemko A, Küzmič F, Landucci F, Lee MT, Levesley A, Li C-F, Liu H, Lopez-Gonzalez G, Lysenko T, Macanović A, Mahdavi P, Manning P, Marcenò C, Martynenko V, Mencuccini M, Minden V, Moeslund JE, Moretti M, Müller JV, Munzinger J, Niinemets Ü, Nobis M, Noroozi J, Nowak A, Onyshchenko V, Overbeck GE, Ozinga WA, Pauchard A, Pedashenko H, Peñuelas J, Pérez-Haase A, Peterka T, Petřík P, Phillips OL, Prokhorov V, Rašomavičius V, Revermann R, Rodwell J, Ruprecht E, Rūsipa S, Samimi C, Schaminée JHJ, Schmiedel U, Šibík J, Šilc U, Škvorec Ž, Smyth A, Sop T, Sopotlieva D, Sparrow B, Stančić Z, Svenning J-C, Swacha G, Tang Z, Tsiripidis I, Turtureanu PD, Ugurlu E, Uogintas D, Valachovič M, Vanselow KA, Vashenyak Y, Vassilev K, Vélez-Martin E, Venanzoni R, Vibrans AC, Violle C, Virtanen R, von Wehrden H, Wagner V, Walker DA, Wana D, Weiher E, Wesche K, Whitfield T, Willner W, Wiser S, Wohlgemuth T, Yamalov S, Zizka G, Zverev A, Chiarucci A (2019) sPlot—a new tool for global vegetation analyses. *J Veg Sci* 30(2):161–186
- Cain SA (1950) Life-forms and phytoclimate. *Bot Rev* 16(1):1–32
- Cavender-Bares J, Gamon JA, Townsend PA (2020) Remote sensing of plant biodiversity. Springer, Cham
- Chicco D (2017) Ten quick tips for machine learning in computational biology. *BioData Min* 10(1):35

- Chytrý M, Schaminée JHJ, Schwabe A (2011) Vegetation survey: a new focus for Applied Vegetation Science. *Appl Veg Sci* 14(4):435–439
- Cook CN, Wardell-Johnson G, Keatley M, Gowans SA, Gibson MS, Westbrooke ME, Marshall DJ (2010) Is what you see what you get? Visual vs. measured assessments of vegetation condition. *J Appl Ecol* 47(3):650–661
- De Cáceres M, Wiser SK (2011) Towards consistency in vegetation classification. *J Veg Sci* 23(2):387–393
- DeFries RS, Field CB, Fung I, Justice CO, Los S, Matson PA, Matthews E, Mooney HA, Potter CS, Prentice K, Sellers PJ, Townshend JRG, Tucker CJ, Ustin SL, Vitousek PM (1995) Mapping the land surface for global atmosphere-biosphere models: toward continuous distributions of vegetation's functional properties. *J Geophys Res: Atmos* 100(D10):20867–20882
- Dengler J, Jansen F, Glöckler F, Peet RK, De Cáceres M, Chytrý M, Ewald J, Oldeland J, Lopez-Gonzalez G, Finckh M, Mucina L, Rodwell JS, Schaminée JHJ, Spencer N (2011) The global index of vegetation-plot databases (GIVD): a new resource for vegetation science. *J Veg Sci* 22(4):582–597
- Dormann CF, McPherson JM, Araújo MB, Bivand R, Bolliger J, Carl G, Davies RG, Hirzel A, Jetz W, Daniel Kissling W, Kühn I, Ohlemüller R, Peres-Neto PR, Reineking B, Schröder B, Schurr FM, Wilson R (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 30(5):609–628
- Drielsma M, Ferrier S (2006) Landscape scenario modelling of vegetation condition. *Ecol Manag Restor* 7(S1):S45–S52
- Elith J, Graham CH, Anderson RP, Dudík M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, McC Overton JM, Townsend Peterson A, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberón J, Williams S, Wisz MS, Zimmermann NE (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29(2):129–151
- Engemann K, Sandel B, Boyle BB, Enquist BJ, Jørgensen PM, Kattge J, McGill BJ, Morueta-Holme N, Peet RK, Spencer NJ, Violle C, Wiser SK, Svenning JC (2016) A plant growth form dataset for the New World. *Ecology* 97(11):3243–3243
- Evans JS, Cushman SA (2009) Gradient modeling of conifer species using random forests. *Landsc Ecol* 24(5):673–683
- Evans MC, Watson JEM, Fuller RA, Venter O, Bennett SC, Marsack PR, Possingham HP (2011) The spatial distribution of threats to species in Australia. *Bioscience* 61(4):281–289
- Ferrier S, Drielsma M (2010) Synthesis of pattern and process in biodiversity conservation assessment: a flexible whole-landscape modelling framework. *Divers Distrib* 16(3):386–402
- Ferrier S, Guisan A (2006) Spatial modelling of biodiversity at the community level. *J Appl Ecol* 43(3):393–404
- Fick SE, Hijmans RJ (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int J Climatol* 37(12):4302–4315
- Fielding AH (1999) Machine learning methods for ecological applications. In: Fielding AH (ed) *Generic*. Springer, Boston
- Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Glob Ecol Biogeogr* 16(3):265–280
- Fleishman E, Noss RF, Noon BR (2006) Utility and limitations of species richness metrics for conservation planning. *Ecol Ind* 6(3):543–553
- Franklin J (2009) Mapping species distributions: spatial inference and prediction. Cambridge University Press, Cambridge
- Franklin J, Serra-Diaz JM, Syphard AD, Regan HM (2017) Big data for forecasting the impacts of global change on plant communities. *Glob Ecol Biogeogr* 26(1):6–17
- Gallant JC, Dowling TI, Read AM, Wilson N, Tickle P, Inskeep C (2011) 1 second SRTM derived digital elevation models user guide, 1.0.4. Geoscience Australia Canberra, Symonston
- Gardner TA, Barlow J, Chazdon RR, Ewers RM, Harvey CA, Peres CA, Sodhi NS (2009) Prospects for tropical forest biodiversity in a human-modified world. *Ecol Lett* 12(6):561–582
- Getis A, Ord JK (2010) The analysis of spatial association by use of distance statistics. In: Anselin L, Ray S (eds) *Perspectives on spatial data analysis. Advances in Spatial Science (The Regional Science Series)*. Springer, Berlin, pp 127–145
- Gilliam FS (2007) The ecological significance of the herbaceous layer in temperate forest ecosystems. *Bioscience* 57(10):845–858
- Grime JP (1977) Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *Am Nat* 111(982):1169–1194
- Grossman DH, Faber-Langendoen D, Weakley A, Anderson M, Bourgeron P, Crawford R, Goodin K, Landaal S, Metzler K, Patterson K (1998) International classification of ecological communities: terrestrial vegetation of the United States. The Nature Conservancy, Arlington
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecol Model* 135(2–3):147–186
- Harden GJ (1990) *Flora of New South Wales*. NSW University Press, Kensington
- Hearn SM, Healey JR, McDonald MA, Turner AJ, Wong JLG, Stewart GB (2011) The repeatability of vegetation classification and mapping. *J Environ Manag* 92(4):1174–1184
- Heikkinen RK, Marmion M, Luoto M (2012) Does the interpolation accuracy of species distribution models come at the expense of transferability? *Ecography* 35(3):276–288
- Hengl T, Mendes de Jesus J, Heuvelink GBM, Ruiperez Gonzalez M, Kilibarda M, Blagotić A, Shangguan W, Wright MN, Geng X, Bauer-Marschallinger B, Guevara MA, Vargas R, MacMillan RA, Batjes NH, Leenaars JGB, Ribeiro E, Wheeler I, Mantel S, Kempen B (2017) SoilGrids250m: global gridded soil information based on machine learning. *PLoS ONE* 12(2):e0169748
- Hooper DU, Chapin FS III, Ewel JJ, Hector A, Inchausti P, Lavorel S, Lawton JH, Lodge DM, Loreau M, Naeem S, Schmid B, Setälä H, Symstad AJ, Vandermeer J, Wardle DA (2005) Effects of biodiversity on ecosystem

- functioning: a consensus of current knowledge. *Ecol Monogr* 75(1):3–35
- Houborg R, Fisher JB, Skidmore AK (2015) Advances in remote sensing of vegetation function and traits. *Int J Appl Earth Obs Geoinf* 43:1–6
- Johnson DP, Driscoll DA, Catford JA, Gibbons P (2020) Fine-scale variables associated with the presence of native forbs in natural temperate grassland. *Austral Ecol* 45(3):366–375
- Kattge J, Bönsch G, Díaz S, Lavorel S, Prentice IC, Leadley P, Tautenhahn S, Werner GDA, Aakala T, Abedi M, Acosta ATR, Adamidis GC, Adamson K, Aiba M, Albert CH, Alcántara JM, Alcázar CC, Aleixo I, Ali H, Amiaud B, Ammer C, Amoroso MM, Anand M, Anderson C, Anten N, Antos J, Apgaua MDG, Ashman T-L, Asmara DH, Asner GP, Aspinwall M, Atkin O, Aubin I, Baastrup-Spohr L, Bahalkeh K, Bahn M, Baker T, Baker WJ, Bakker JP, Baldocchi D, Baltzer J, Banerjee A, Baranger A, Barlow J, Barneche DR, Baruch Z, Bastianelli D, Battles J, Bauerle W, Bauters M, Bazzato E, Beckmann M, Biecekman H, Beierkuhnlein C, Bekker R, Belfry G, Belluau M, Beloiu M, Benavides R, Benomar L, Berdugo-Lattke ML, Berenguer E, Bergamin R, Bergmann J, Bergmann Carlucci M, Berner L, Bernhardt-Römermann M, Bigler C, Bjorkman AD, Blackman C, Blanco C, Blonder B, Blumenthal D, Bocanegra-González KT, Boeckx P, Bohlman S, Böhning-Gaese K, Boisvert-Marsh L, Bond W, Bond-Lamberty B, Boom A, Boonman CCF, Bordin K, Boughton EH, Boukili V, Bowman DMJS, Bravo S, Brendel MR, Broadley MR, Brown KA, Bruelheide H, Brummich F, Bruun HH, Bruy D, Buchanan SW, Bucher SF, Buchmann N, Buitenwerf R, Bunker DE, Bürger J, Burrascano S, Burslem DFRP, Butterfield BJ, Byun C, Marques M, Scalón MC, Caccianiga M, Cadotte M, Cailleret M, Camac J, Camarero JJ, Company C, Campetella G, Campos JA, Cano-Arboleda L, Canullo R, Carbognani M, Carvalho F, Casanoves F, Castagnayrol B, Catford JA, Cavender-Bares J, Cerabolini BEL, Cervellini M, Chacón-Madriral E, Chapin K, Chapin FS, Chelli S, Chen S-C, Chen A, Cherubini P, Chianucci F, Choat B, Chung K-S, Chytrý M, Ciccarelli D, Coll L, Collins CG, Conti L, Coomes D, Cornelissen JHC, Cornwell WK, Corona P, Coyea M, Craine J, Craven D, Croomsig JPM, Csécsérts A, Cufar K, Cuntz M, da Silva AC, Dahlin KM, Dainese M, Dalke I, Dalle Fratte M, Dang-Le AT, Danihelka J, Dannoura M, Dawson S, de Beer AJ, De Frutos A, De Long JR, Dechant B, Delagrange S, Delpierre N, Derroire G, Dias AS, Diaz-Toribio MH, Dimitrakopoulos PG, Dobrowolski M, Doktor D, Dřevojan P, Dong N, Dransfield J, Dressler S, Duarte L, Ducouret E, Dullinger S, Durka W, Duursma R, Dymova O, E-Vojtkó A, Eckstein RL, Ejtehadi H, Elser J, Emilio T, Engemann K, Erfanian MB, Erfmeier A, Esquivel-Muelbert A, Esser G, Estiarte M, Domingues TF, Fagan WF, Fagúndez J, Falster DS, Fan Y, Fang J, Farris E, Fazlioglu F, Feng Y, Fernandez-Mendez F, Ferrara C, Ferreira J, Fidelis A, Finegan B, Firn J, Flowers TJ, Flynn DFB, Fontana V, Forey E, Forgiarini C, François L, Frangipani M, Frank D, Frenette-Dussault C, Freschet GT, Fry EL, Fyllas NM, Mazzochini GG, Gachet S, Gallagher R, Ganade G, Ganga F, García-Palacios P, Gargaglione V, Garnier E, Garrido JL, de Gasper AL, Gea-Izquierdo G, Gibson D, Gillison AN, Girollo A, Glasenhardt M-C, Gleason S, Gliesch M, Goldberg E, Göddel B, Gonzalez-Akre E, Gonzalez-Andujar JL, González-Melo A, González-Robles A, Graae BJ, Granda E, Graves S, Green WA, Gregor T, Gross N, Guerin GR, Günther A, Gutiérrez AG, Haddock L, Haines A, Hall J, Hambuckers A, Han W, Harrison SP, Hattening W, Hawes JE, He T, He P, Heberling JM, Helm A, Hempel S, Hentschel J, Hérault B, Hereş A-M, Herz K, Heuertz M, Hickler T, Hietz P, Higuchi P, Hipp AL, Hiron A, Hock M, Hogan JA, Holl K, Honnay O, Hornstein D, Hou E, Hough-Snee N, Hovstad KA, Ichie T, Igić B, Illa E, Isaac M, Ishihara M, Ivanov L, Ivanova L, Iversen CM, Izquierdo J, Jackson RB, Jackson B, Jactel H, Jagodzinski AM, Jandt U, Jansen S, Jenkins T, Jentsch A, Jespersen JRP, Jiang G-F, Johansen JL, Johnson D, Jokela EJ, Joly CA, Jordan GJ, Joseph GS, Junaedi D, Junker RR, Justes E, Kabzems R, Kane J, Kaplan Z, Kattenborn T, Kavelenova L, Kearsley E, Kempel A, Kenzo T, Kerckhoff A, Khalil MI, Kinlock NL, Kissling WD, Kitajima K, Kitzberger T, Kjoller R, Klein T, Kleyer M, Klimešová J, Klipel J, Kloppel B, Klotz S, Knops JMH, Kohyama T, Koike F, Kollmann J, Komac B, Komatsu K, König C, Kraft NJB, Kramer K, Kreft H, Kühn I, Kumarathunge D, Kuppler J, Kurokawa H, Kurosawa Y, Kuyah S, Laclau J-P, Lafleur B, Lallai E, Lamb E, Lamprecht A, Larkin DJ, Laughlin D, Le Bagousse-Pinguet Y, le Maire G, le Roux PC, le Roux E, Lee T, Lens F, Lewis SL, Lhotsky B, Li Y, Li X, Lichstein JW, Liebergesell M, Lim JY, Lin Y-S, Linares JC, Liu C, Liu D, Liu U, Livingstone S, Llusà J, Lohbeck M, López-García Á, Lopez-Gonzalez G, Lososová Z, Louault F, Lukács BA, Lukeš P, Luo Y, Lussu M, Ma S, Maciel Rabelo Pereira C, Mack M, Maire V, Mäkelä A, Mäkinen H, Malhado ACM, Mallik A, Manning P, Manzoni S, Marchetti Z, Marchino L, Marcilio-Silva V, Marcon E, Marignani M, Markesteijn L, Martin A, Martínez-Garza C, Martínez-Vilalta J, Mašková T, Mason K, Mason N, Massad TJ, Masse J, Mayrose I, McCarthy J, McCormack ML, McCulloh K, McFadden IR, McGill BJ, McPartland MY, Medeiros JS, Medlyn B, Meerts P, Mehrabi Z, Meir P, Melo FPL, Mencuccini M, Meredieu C, Messier J, Mészáros I, Metsaranta J, Michaletz ST, Michelaki C, Migalina S, Milla R, Miller JED, Minden V, Ming R, Mokany K, Moles AT, Molnár V A, Molofsky J, Molz M, Montgomery RA, Monty A, Moravcová L, Moreno-Martínez A, Moretti M, Mori AS, Mori S, Morris D, Morrison J, Mucina L, Mueller S, Muir CD, Müller SC, Munoz F, Myers-Smith IH, Myster RW, Nagano M, Naidu S, Narayanan A, Natesan B, Negoita L, Nelson AS, Neuschulz EL, Ni J, Niedrist G, Nieto J, Niinemets Ü, Nolan R, Nottebrock H, Nouvellon Y, Novakovskiy A, Network TN, Nystuen KO, O'Grady A, O'Hara K, O'Reilly-Nugent A, Oakley S, Oberhuber W, Ohtsuka T, Oliveira R, Öllerer K, Olson ME, Onipchenko V, Onoda Y, Onstein RE, Ordóñez JC, Osada N, Ostonen I, Ottaviani G, Otto S, Overbeck GE, Ozinga WA, Pahl AT, Paine CET, Pakeman RJ, Papa-georgiou AC, Parfionova E, Pärtel M, Patacca M, Paula S, Paule J, Pauli H, Pausas JG, Peco B, Penuelas J, Perea A, Peri PL, Petisco-Souza AC, Petraglia A, Petritan AM, Phillips OL, Pierce S, Pillar VD, Pisek J, Pomogaybin A, Poorter H, Portsmouth A, Poschod P, Potvin C, Pounds D, Powell AS, Power SA, Prinzing A, Puglielli G, Pyšek P, Ravel V, Rammig A, Ransijn J, Ray CA, Reich PB,

- Reichstein M, Reid DEB, Réjou-Méchain M, de Dios VR, Ribeiro S, Richardson S, Riibak K, Rillig MC, Riviera F, Robert EMR, Roberts S, Robroek B, Roddy A, Rodrigues AV, Rogers A, Rollinson E, Rolo V, Römermann C, Ronzhina D, Roscher C, Rosell JA, Rosenfield MF, Rossi C, Roy DB, Royer-Tardif S, Rüger N, Ruiz-Peinado R, Rumpf SB, Rusch GM, Ryo M, Sack L, Saldaña A, Salgado-Negret B, Salguero-Gomez R, Santa-Regina I, Santacruz-García AC, Santos J, Sardans J, Schamp B, Scherer-Lorenzen M, Schleuning M, Schmid B, Schmidt M, Schmitt S, Schneider JV, Schowanek SD, Schrader J, Schrodt F, Schuldt B, Schurr F, Selaya Garvizu G, Semchenko M, Seymour C, Sfair JC, Sharpe JM, Sheppard CS, Sheremetiev S, Shiodera S, Shipley B, Shovon TA, Siebenkäs A, Sierra C, Silva V, Silva M, Sitzia T, Sjöman H, Slot M, Smith NG, Sodhi D, Soltis P, Soltis D, Somers B, Sonnier G, Sørensen MV, Sosinski Jr EE, Soudzilovskaia NA, Souza AF, Spasojevic M, Sperandii MG, Stan AB, Stegen J, Steinbauer K, Stephan JG, Sterck F, Stojanovic DB, Strydom T, Suarez ML, Svenning J-C, Svitková I, Svitok M, Svoboda M, Swaine E, Swenson N, Tabarelli M, Takagi K, Tappeiner U, Tarifa R, Tauougrdeau S, Tavsanoğlu C, te Beest M, Tedersoo L, Thiffault N, Thom D, Thomas E, Thompson K, Thornton PE, Thuiller W, Tichý L, Tissue D, Tjoelker MG, Tng DYP, Tobias J, Török P, Tarin T, Torres-Ruiz JM, Tóthmérész B, Treurnicht M, Trivellone V, Trolliet F, Trotsiuk V, Tsakalos JL, Tsiripidis I, Tyskland N, Umehara T, Usoltsev V, Vadeboncoeur M, Vaezi J, Valladares F, Vamasi J, van Bodegom PM, van Breugel M, Van Cleemput E, van de Weg M, van der Merwe S, van der Plas F, van der Sande MT, van Kleunen M, Van Meerbeek K, Vanderwel M, Vanselow KA, Vårhammar A, Varone L, Vasquez Valderrama MY, Vassilev K, Vellend M, Veneklaas EJ, Verbeek H, Verheyen K, Vibrans A, Vieira I, Villacís J, Violle C, Vivek P, Wagner K, Waldram M, Waldron A, Walker AP, Waller M, Walther G, Wang H, Wang F, Wang W, Watkins H, Watkins J, Weber U, Weedon JT, Wei L, Weigelt P, Weiher E, Wells AW, Wellstein C, Wenk E, Westoby M, Westwood A, White PJ, Whitten M, Williams M, Winkler DE, Winter K, Womack C, Wright IJ, Wright SJ, Wright J, Pinho BX, Ximenes F, Yamada T, Yamaji K, Yanai R, Yankov N, Yguel B, Zanini KJ, Zanne AE, Zelený D, Zhao Y-P, Zheng J, Zheng J, Ziemińska K, Zirbel CR, Zizka G, Zo-Bi IC, Zotz G, Wirth C, (2020) TRY plant trait database—enhanced coverage and open access. *Glob Change Biol* 26(1):119–188
- Keith D (2004) Ocean shores to desert dunes: the native vegetation of New South Wales and the ACT. Department of Environment and Conservation NSW, Sydney
- Keitt TH, Bjørnstad ON, Dixon PM, Citron-Pousty S (2002) Accounting for spatial pattern when modeling organism-environment interactions. *Ecography* 25(5):616–625
- Kelemen A, Török P, Valkó O, Migléc T, Tóthmérész B (2013) Mechanisms shaping plant biomass and species richness: plant strategies and litter effect in alkali and loess grasslands. *J Veg Sci* 24(6):1195–1203
- Kissling WD, Ahumada JA, Bowser A, Fernandez M, Fernández N, García EA, Guralnick RP, Isaac NJB, Kelling S, Los W, McRae L, Mihoub J-B, Obst M, Santamaria M, Skidmore AK, Williams KJ, Agosti D, Amariles D, Arvanitidis C, Bastin L, De Leo F, Egloff W, Elith J, Hobern D, Martin D, Pereira HM, Pesole G, Peterseil J, Saarenmaa H, Schigel D, Schmeller DS, Segata N, Turak E, Uhlir PF, Wee B, Hardisty AR (2018) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biol Rev* 93(1):600–625
- Lake PS (2000) Disturbance, patchiness, and diversity in streams. *J N Am Benthol Soc* 19(4):573–592
- Lausch A, Heurich M, Magdon PP, Rocchini D, Schulz K, Bumberger J, King DJ (2020) A range of earth observation techniques for assessing plant diversity. In: Cavender-Bares J, Gamon JA, Townsend PA (eds) Remote sensing of plant biodiversity. Springer, Cham, pp 309–348
- Leitão PJ, Santos MJ (2019) Improving models of species ecological niches: a remote sensing overview. *Front Ecol Evol* 7(9):1–7
- Lesslie R, Barson M, Smith J (2006) Land use information for integrated natural resources management—a coordinated national mapping program for Australia. *J Land Use Sci* 1(1):45–62
- Levin SA (1992) The problem of pattern and scale in ecology. *Ecology* 73(6):1943–1967
- Lindenmayer DB, Blanchard W, Crane M, Michael D, Florance D (2018) Size or quality. What matters in vegetation restoration for bird biodiversity in endangered temperate woodlands? *Austral Ecol* 43(7):798–806
- Lucas RM, Cronin N, Moghaddam M, Lee A, Armston J, Bunting P, Witte C (2006) Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. *Remote Sens Environ* 100(3):388–406
- Lunt ID, Eldridge DJ, Morgan JW, Witt GB (2007) A framework to predict the effects of livestock grazing and grazing exclusion on conservation values in natural ecosystems in Australia. *Aust J Bot* 55(4):401–415
- Margules CR, Pressey RL (2000) Systematic conservation planning. *Nature* 405(6783):243–253
- Maxwell SL, Fuller RA, Brooks TM, Watson JE (2016) Biodiversity: the ravages of guns, nets and bulldozers. *Nat News* 536(7615):143
- McElhinny C, Gibbons P, Brack C, Bauhus J (2005) Forest and woodland stand structural complexity: its definition and measurement. *For Ecol Manag* 218(1–3):1–24
- McElhinny C, Gibbons P, Brack C, Bauhus J (2006) Fauna-habitat relationships: a basis for identifying key stand structural attributes in temperate Australian eucalypt forests and woodlands. *Pac Conserv Biol* 12(2):89–110
- McNellie MJ, Oliver I, Gibbons P (2015) Pitfalls and possible solutions for using geo-referenced site data to inform vegetation models. *Ecol Inform* 30:230–234
- McNellie MJ, Dorrrough J, Oliver I (2019) Species abundance distributions should underpin ordinal cover-abundance transformations. *Appl Veg Sci* 22(3):361–372
- Miller J (2005) Incorporating spatial dependence in predictive vegetation models: residual interpolation methods. *Prof Geogr* 57(2):169–184
- Morrison LW (2016) Observer error in vegetation surveys: a review. *J Plant Ecol* 9(4):367–379
- Mucina L, van der Maarel E (1989) Twenty years of numerical syntaxonomy. In: Mucina L, Dale MB (eds) Numerical syntaxonomy. Advances in vegetation science. Springer, Dordrecht

- Newbold T, Hudson LN, Arnell AP, Contu S, De Palma A, Ferrier S, Hill SLL, Hoskins AJ, Lysenko I, Phillips HRP, Burton VJ, Chng CWT, Emerson S, Gao D, Pask-Hale G, Hutton J, Jung M, Sanchez-Ortiz K, Simmons BI, Whitmee S, Zhang H, Scharlemann JPW, Purvis A (2016) Has land use pushed terrestrial biodiversity beyond the planetary boundary? A global assessment. *Science* 353(6296):288
- Noss RF (1990) Indicators for monitoring biodiversity: a hierarchical approach. *Conserv Biol* 4(4):355–364
- NSW Office of Environment and Heritage (2017) The NSW State Vegetation Type Map: methodology for a regional-scale map of NSW plant community types. In: NSW Office of Environment and Heritage (ed) Sydney, Australia
- O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. *Qual Quant* 41(5):673–690
- Olf H, Ritchie ME (1998) Effects of herbivores on grassland plant diversity. *Trends Ecol Evol* 13(7):261–265
- Oliver I, McNellie MJ, Steenbeeke G, Copeland L, Porteners MF, Wall J (2019) Expert allocation of primary growth form to the NSW flora underpins the Biodiversity Assessment Method. *Australas J Environ Manag* 26(2):124–136
- Oliver I, Dorrrough J, Seidel J (2021) A new vegetation integrity metric for trading losses and gains in terrestrial biodiversity value. *Ecol Ind* 124:107341
- Opitz D, Maclin R (1999) Popular ensemble methods: an empirical study. *J Artif Intell Res* 11:169–198
- Paramasivam CR, Venkatramanan S (2019) An introduction to various spatial analysis techniques. In: Venkatramanan S, Prasanna MV, Chung SY (eds) GIS and geostatistical techniques for groundwater science. Elsevier, Amsterdam, pp 23–30
- Pausas JG, Austin MP (2001) Patterns of plant species richness in relation to different environments: an appraisal. *J Veg Sci* 12(2):153–166
- Pausas JG, Bond WJ (2019) Humboldt and the reinvention of nature. *J Ecol* 107(3):1031–1037
- Peet RK, Lee MT, Jennings MD, Faber-Langendoen D (2013) VegBank: the vegetation plot archive of the Ecological Society of America. <http://vegbank.org>
- Pressey RL, Hager TC, Ryan KM, Schwarz J, Wall S, Ferrier S, Creaser PM (2000) Using abiotic data for conservation assessments over extensive regions: quantitative methods applied across New South Wales. Australia. *Biol Conserv* 96(1):55–82
- Pressey RL, Cabeza M, Watts ME, Cowling RM, Wilson KA (2007) Conservation planning in a changing world. *Trends Ecol Evol* 22(11):583–592
- Quinn TP, Le V, Cardilini APA (2021) Test set verification is an essential step in model building. *Methods Ecol Evol* 12(1):127–129
- Rowe N, Speck T (2005) Plant growth forms: an ecological and evolutionary perspective. *New Phytol* 166(1):61–72
- Sajid A, Rudra R, Parkin G (2013) Systematic evaluation of kriging and inverse distance weighting methods for spatial analysis of soil bulk density. *Can Biosyst Eng* 55(1):1–13
- Saunders DA, Hobbs RJ, Margules CR (1991) Biological consequences of ecosystem fragmentation: a review. *Conserv Biol* 5(1):18–32
- Scarath P, Röder A, Schmidt M, Denham R (2010) Tracking grazing pressure and climate interaction-the role of Landsat fractional cover in time series analysis. In: Proceedings of the 15th Australasian remote sensing and photogrammetry conference, Alice Springs, Australia, 2010. vol 13. Published by the Remote Sensing and Photogrammetry Commission
- Schaminée JHJ, Janssen JAM, Hennekens SM, Ozinga WA (2011) Large vegetation databases and information systems: new instruments for ecological research, nature conservation, and policy making. *Plant Biosyst* 145(sup1):85–90
- Schimel D, Townsend PA, Pavlick R (2020) Prospects and pitfalls for spectroscopic remote sensing of biodiversity at the global scale. In: Cavender-Bares J, Gamon JA, Townsend PA (eds) Remote sensing of plant biodiversity. Springer, Cham, pp 503–518
- Schrodt F, de la Barreda Bautista B, Williams C, Boyd DS, Schaepman-Strub G, Santos MJ (2020) Integrating biodiversity, remote sensing, and auxiliary information for the study of ecosystem functioning and conservation at large spatial scales. In: Cavender-Bares J, Gamon JA, Townsend PA (eds) Remote sensing of plant biodiversity. Springer, Cham, pp 449–484
- Shmueli G (2010) To explain or to predict? *Stat Sci* 25(3):289–310
- Smith WG (1913) Raunkiaer's "Life-Forms" and statistical methods. *J Ecol* 1(1):16–26
- Speed JDM, Austrheim G (2017) The importance of herbivore density and management as determinants of the distribution of rare plant species. *Biol Conserv* 205:77–84
- Statsoft Inc. (2013) Electronic statistics textbook. Tulsa, Oklahoma. <http://www.statsoft.com/textbook/>
- Svenning JC, Sandel B (2013) Disequilibrium vegetation dynamics under future climate change. *Am J Bot* 100(7):1266–1286
- Taylor C, Lindenmayer DB (2020) Temporal fragmentation of a critically endangered forest ecosystem. *Austral Ecol* 45(3):340–354
- Turner MG (2005) Landscape ecology in North America: past, present, and future. *Ecology* 86(8):1967–1974
- Ustin SL, Gamon JA (2010) Remote sensing of plant functional types. *New Phytol* 186(4):795–816
- Warming E, Vahl M (1909) *Oecology of plants*. Oxford University Press, London
- Williams KJ, Belbin L, Austin MP, Stein JL, Ferrier S (2012) Which environmental variables should I use in my biodiversity model? *Int J Geogr Inf Sci* 26(11):2009–2047
- Wintle BA, Elith J, Potts JM (2005) Fauna habitat modelling and mapping: a review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecol* 30(7):719–738
- Wiser SK, Hurst JM, Wright EF, Allen RB (2011) New Zealand's forest and shrubland communities: a quantitative classification based on a nationally representative plot network. *Appl Veg Sci* 14(4):506–523
- Xu T, Hutchinson MF (2013) New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. *Environ Modell Softw* 40:267–279