Supplementary Material

This supplementary material provides details behind the conceptual underpinning of selecting predictor variables, model architecture, software tools and mapped vegetation attributes with residual errors for the paper McNellie, M.J., Oliver, I., Ferrier, S., Newell, G., Manion, G., Griffioen, P., White, M., Somerville, M., Koen, T. and Gibbons, P. ***Extending vegetation site data and ensemble models to predict regional patterns of cover and richness for plant functional groups*.**

Table of Contents

Supplementary Material S1 – Conceptual underpinning of the environmental and disturbance gradients – the predictor variables 2

Supplementary Material S2 – Testing for correlated and collinear predictor variables 12

Supplementary Material S3 – Additional information on artificial neural network model architecture 15

Supplementary Material S4 – Software to create data input tables, prediction output tables and manage multiple tables 19

Supplementary Material S5 – Equations for calculating root mean squared deviation (RMSD) and mean absolute error (MAE) 21

Supplementary Material S6 – Sensitivity analysis showing the sensitivity averaged over 25 ensembles of ANN models for ten vegetation attributes 22

Supplementary Material S7 – Results to assess for spatial autocorrelation - Moran's Index 23

Supplementary Material S8 – The spatially explicit maps and accompanying residual maps for each vegetation attribute 25

Supplementary Material S1 – Conceptual underpinning of the environmental and disturbance gradients – the predictor variables

We selected a suite of predictor surfaces based on our ecological understanding of the landscape and conceptual understanding of drivers that influence patterns in vegetation; 'V' indicates that the specified predictor surface is known to influence the growth and morphology of vegetation extent or type; 'M' indicates surfaces included to inform modification or fragmentation of vegetation. 'V,M' indicates that within a whole of landscape context, the predictor informs both extent and type of vegetation as well as modification and fragmentation.

Climate surfaces were generated using MTHCLIM in ANUCLIM v6.1 (Xu and Hutchinson 2013) for the 1921 -1995 epoch. A 25 m resolution 'smoothed' digital elevation model DEM-S (Gallant et al. 2011) was used as the third independent variable needed to interpolate monthly mean climate values. These variables were selected to express various aspects of the annual regimes of temperature, precipitation, and potential evapotranspiration (Box 1981). After ANUCLIM processing, climate surfaces were imported into ArcGIS v10.1, converted to raster layers (25 m) and projected to Lamberts NSW coordinate system. Sharples et al. (2005) note that digital elevation models with a spatial scale of 1 km resolution are normally sufficient for temperature dependent parameters and a 5 km resolution DEM is normally sufficient for rainfall dependent parameters. Table S1 outlines the predictor surfaces used to represent the environmental and disturbance gradients.

Table S1 Suite of potential predictor surfaces identified as ecologically realistic variables used in artificial neural network models.

Environmental gradient	Predictor surface	Informs	Description	Response Data Range	Conceptual underpinning or reference to method
Macroclimate	isothermality	V	temperature evenness as calculated in BIOCLIM (BIO03) as the diurnal temperature (°C) range divided by annual temperature range. It is a measure of the day to night temperature oscillation in comparison to the seasonal oscillation. annual range of monthly mean temperatures	0.43 – 0.52	Xu & Hutchinson 2013
	mean minimum temp in winter	V	averaged minimum temperature (°C) for June, July and August	-1.3 – 5.2	Box 1981
	winter rainfall	V	averaged rainfall (mm) for June, July and August	78 – 334	Box 1981
	mean maximum temperature in summer	V	averaged maximum temperature (°C) December, January and February	20.5 – 34.3	Box 1981
	summer rainfall	V	total rainfall (mm) for December, January and February	137.8 – 690.5	Box 1981
	rainfall seasonality ratio	V	solstice rainfall seasonality ratio calculated as rainfall (mm) totals for December, January and February over rainfall totals for June, July and August	1.1 – 2.9	Austin 1998

	precipitation/ evaporation	V	annual moisture index (annual precipitation divided by the estimate of annual potential evapotranspiration)	0.17 – 1.7	Box 1981
	heat load index	V	Input variables are latitude, slope, and aspect. The equations apply to 0–60° north latitude, slopes from 0–90°, and all aspects. By transforming aspect, the equations can also be applied as an index of heat load, symmetrical about a northeast to southwest axis.	1025 - 7088	McCune & Keon 2002
Topography	transformed aspect	V	Transformed aspect = 1 - cos (aspect - 30)/2. This transformation assigns the highest values to land oriented in a south-southwest direction, the coolest and wettest orientation for Australian landscapes (southern hemisphere).	0 – 1	Roberts & Cooper 1989
	slope	V	identifies the rate of maximum change in elevation between each cell	0 – 52	
	compound topographic index	V	Generated from the hydrologically enforced digital elevation model (DEM-H see Gallant et al. 2011)	3.5 – 20.0	Gallant et al. 2011; Gessler et al. 1995
	Distance to drainage line	V	continuous surfaces that represent Euclidian distance (m) to the nearest drainage line		
	Distance to wetland	V	continuous surfaces that represent Euclidian distance (m) to the nearest wetland		

Soils and Geology	gravity	V	Anomalies in the earth's gravitational field can in some cases be used to define the thickness and extent of the regolith, due to the strong contrast in density between the regolith (including in situ weathered bedrock) and fresh bedrock, and in defining regolith features such as paleo-channels and is used *in lieu* of geology	-267 – 442	Tracey et al. 2007; Wellman 1998
Radiometric	ratio Th/K		Gamma radiometric data (K, U and Th) provides information on the distribution of materials in which the soils. Ratios of these gamma-emissions show the spatial distribution of a range of contrasting soil-forming materials. Potassium (K) (%) concentration decreases with increasing weathering and occurs in many rock-forming minerals such as K-feldspars, micas and as clays like illite. Thorium (Th) (ppm) is most common in accessory and resistate minerals such as zircon, sphene, apatite, xenotime, monazite and epidote.	0.15 – 218.	Cook et al. 1996
Soils	great soil group	V,M	Mapped great soil groups were ordered according to their profile development.	41 categories	Murphy 2000
	% clay, %sand and %silt		The clay, silt and sand content of a soil control its texture and physical behaviour, including water		OEH 2017; Viscarra Rossel 2011.

			holding capacity and permeability, and influence many chemical characteristics. Clay content in soils influences a wide range of physical processes related to hydrology, mineral status and fertility		

Environmental Gradient	Predictor Surface	Informs		Range	Conceptual underpinning or reference to method
Landsat Imagery	median band 3: median band 4: median band 5: median band 7	V,M	Median pixel value for each Landsat image sliced into individual bands 3, 4, 5 and 7 for years 2005-2012.	10 – 174 6 – 169 7 – 227 6 – 124	
	foliage projected cover	V,M	Estimated foliage cover for woody areas derived from Landsat imagery 2005 - 2012. Foliage projected cover and fractional groundcover are mutually exclusive data sets. The fractional ground cover has only been assessed in areas with less than 20% foliage cover.	100 - 199	Lucas et al. 2006.
	fractional groundcover index green groundcover: brown groundcover:	V,M	A Landsat 7 ground cover image based on a calibrated relationship between field measurement and the imagery from 2005 – 2012. These data are	 87 – 201 87 – 216	McCosker et al. 2009 Scarth et al. 2010

	bare ground		generally used in a time-series analysis for monitoring trends in ground cover over time. The proportions of green (living photosynthetic plant cover) vegetation; brown, non-photosynthetic vegetation (litter cover), and bare (soil, rock) are generated using the algorithm developed by the Joint Remote Sensing Research Program (JRSRP) and described in Scarth et al. (2010). Details for the algorithms to calculate FPC and ground cover are detailed in http://data.auscover.org.au/xwiki/bin/view/Product+pages/nsw+5m+woody+extent+and+fpc and http://data.auscover.org.au/xwiki/bin/view/Product+pages/Seasonal+Ground+Cover, respectively. These algorithms were tailored to the 2005-2012 temporal window.	84 – 202	
	normalised difference vegetation index	V,M	(Band 4 – Band 3) / (Band 4 + Band 3) used to detect vegetation changes in the biophysical parameters of the canopy.	-0.45 – 0.74	
Modification	major land use and landforms	M	7 major broad categories representing proxies for levels disturbance were identified and mapped from aerial photograph interpretation: Percentage of the study area cover by each land use in brackets:	7 categories	Foley et al. 2005; Fischer & Lindenmayer 2007.

			• grazing (52%); • cropping and horticulture (23%); • conservation areas (10%); • tree and shrub cover (10%); • other land uses (which includes all urban areas, roads, mining, power generation and areas used for intensive animal production – 2%); • rivers (2%) and • wetlands (less than 1%). https://datasets.seed.nsw.gov.au/dataset/nsw-landuse-2013/resource/62528ce5-9e6b-4285-8b7c-b1acef014398)		
Location	Eastings and Northings		Including the Easting and Northing coordinates as predictors alongside other environmental variables is a powerful way of addressing spatial autocorrelation and can be useful at broader spatial scales to detect spatial gradients across the study area.		Legendre 1993.

S1 Supplementary References

Austin MP (1998) An ecological perspective on biodiversity investigations: examples from Australian eucalypt forests. Annals of the Missouri Botanical Garden 85(1):2-17 https://doi.org/10.2307/2991991

Box EO (1981) Predicting physiognomic vegetation types with climate variables. Vegetatio 45(2):127-139 https://doi.org/10.2307/20037033

Cook S, Corner R, Groves P, Grealish G (1996) Use of airborne gamma radiometric data for soil mapping. Soil Research 34(1):183-194 https://doi.org/10.1071/SR9960183

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. Global Ecology and Biogeography 16(3):265-280 https://doi.org/10.1111/j.1466-8238.2007.00287.x

Foley JA, DeFries R, Asner GP et al. (2005) Global Consequences of Land Use. Science 309(5734):570 https://doi.org/10.1126/science.1111772

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 edn. Geoscience Australia www.ga.gov.au/topographic-mapping/digital-elevation-data.html, Canberra.

Gessler PE, Moore ID, McKenzie NJ, Ryan PJ (1995) Soil-landscape modelling and spatial prediction of soil attributes. International Journal of Geographical Information Systems 9(4):421-432 https://doi.org/10.1080/02693799508902047

Legendre P (1993) Spatial Autocorrelation: Trouble or New Paradigm? Ecology 74(6):1659-1673 https://doi.org/10.2307/1939924

Lucas RM, Cronin N, Moghaddam M et al. (2006) Integration of radar and Landsat-derived foliage projected cover for woody regrowth mapping, Queensland, Australia. Remote Sensing of Environment 100(3):388-406 https://doi.org/10.1016/j.rse.2005.09.020

McCosker J, Rolfe J, Fensham R (2009) Can bare ground cover server as a surrogate for plant biodiversity in grazed tropical woodlands? The Rangeland Journal 31(1):103-109 https://doi.org/10.1071/rj08041

McCune B, Keon D (2002) Equations for potential annual direct incident radiation and heat load. Journal of Vegetation Science 13(4):603-606 https://doi.org/10.1111/j.1654-1103.2002.tb02087.x

Murphy, B. (2000). Systems of soil classification. In ‘Soils: their properties and management’. 2nd edn (Eds PEV Charman, BW Murphy) pp. 83–105. Oxford University Press: Melbourne.

OEH (2017). Digital soil mapping of key soil properties over NSW, Technical Report, NSW https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Land-and-soil/digital-soil-mapping-key-soil-properties-nsw-170255.pdf

Roberts DW, Cooper SV (1989). Concepts and techniques of vegetation mapping. General technical report pp. 90-96. INT-US Department of Agriculture, Forest Service, Intermountain Research Station.

Scarth 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 (ARSPC), 13-17 September, Alice Springs, Australia. Alice Springs, NT.

Sharples JJ, Hutchinson MF, Jellett DR (2005) On the horizontal scale of elevation dependence of Australian monthly precipitation. Journal of Applied Meteorology 44(12):1850-1865 https://doi.org/10.1175/JAM2289.1

Tracey R, Bacchin M, Wynne P (2007) AAGD07: A new absolute gravity datum for Australian gravity and new standards for the Australian National Gravity Database. ASEG Extended Abstracts 2007(1):1-3 https://doi.org/10.1071/ASEG2007ab149

Viscarra Rossel RA (2011) Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. Journal of Geophysical Research: Earth Surface 116(F4)https://doi.org/10.1029/2011jf001977

Wellman P (1998) Mapping of geophysical domains in the Australian continental crust using gravity and magnetic anomalies. Structure and Evolution of the Australian Continent, Geodyn. Ser. AGU, Washington, DC, pp. 59-71

Xu T, Hutchinson MF (2013) New developments and applications in the ANUCLIM spatial climatic and bioclimatic modelling package. Environ. Modell. Softw. 40(0):267-279 https://doi.org/10.1016/j.envsoft.2012.10.003

Supplementary Material S2 – Testing for correlated and collinear predictor variables

Our choice of potential predictors shown in Supplementary Material S1 was guided by ecologically informed expert-judgement. In total, we prepared 29 potential surfaces that described patterns in macroclimate, soils and geology, topography, landscape modification and remote sensing. Among these predictors, we tested if some were correlated or collinear.

We used Pearson's correlation coefficient to identify which predictors were highly correlated ($r \geq 0.7$ or $r \leq -0.7$). Pearson's was deemed appropriate after visual inspection of scatter plots revealed no strong evidence of curvilinear relationships among variables. Land use and greater soil group were not included in the correlation matrix because they are categorical predictors. Where continuous predictors were highly correlated, one was excluded. We found %clay and %sand were highly correlated (r = 0.84). Seasonal temperature and rainfall predictors were frequently correlated with other climatic predictors (r = 0.75 to 0.94), so we excluded four surfaces that describe temperature and rainfall extremes (mean minimum winter temperature; mean maximum summer temperature; winter rainfall and summer rainfall) and retained three predictor surfaces (isothermality; precipitation/evaporation and rainfall seasonality ratio). We also excluded heat load index (HLI) as it is based on aspect and slope which were included as predictors in the models. We opted to use Normalised Difference Vegetation Index (NDVI) and bare ground surfaces, and exclude the individual Landsat surfaces (median values for band 3; band 4; band 5 and 7) and green groundcover and brown ground cover.

Pearson's correlation matrix is limited because it can only establish the relationship between two predictor variables at a time. In contrast, the Variance Inflation Ratio (VIF) is a practical approach for assessing multicollinearity amongst all predictor variables (Vu et al. 2015), including categorical predictors like land use and soil type. In addition, the VIF calculations are straightforward and comprehensive; the higher the value of the VIF, the higher the collinearity is between the variables. Where the VIF is close to 1, it indicates an absence of collinearity and the VIF of 10 indicates that (all other things being equal) the variance of the i^{th} regression coefficient is 10 times greater than it would have been if the i^{th} independent variable had been linearly independent of the other independent variable in the analysis (O'Brien 2007). The threshold value used to diagnose multicollinearity is a 'rule of thumb' (O'Brien 2007), and can range between 5 (e.g., Citakoglu 2017, Vu et al. 2015) and 10 (Alin 2010).

Here we used the conservative value of >=5 to screen for multicollinearity in the 18 predictors surfaces that were retained after we reviewed the results of Pearson's correlation coefficients. Where the VIF exceeded 5, the predictor was excluded and the VIF was recalculated (O'Brien 2007). Table S2 shows the VIF using the initial suite of 18 predictor surfaces and the revised VIF once Eastings was excluded. Once Eastings (VIF = 12.5) were excluded, the VIF for the remaining 16 predictors did not exceed 3.85.

Table S2 Results of Variance Inflation Factors (VIF) to assess multicollinearity. Numbers in bold show which predictors had a VIF >=5.

Environmental gradient	Original suite of predictor surfaces	VIF	Revised suite of predictor surface	VIF
Macroclimate	isothermality	4.20	isothermality	1.68
	rainfall seasonality ratio	2.03	rainfall seasonality ratio	1.61
	precipitation/evaporation	**7.51**	precipitation/evaporation	2.22
Soil and Geology	gravity	1.85	gravity	1.46
	% clay	2.79	% clay	2.60
	greater soil group	1.17	greater soil group	1.15
Radiometric	Th/K	1.29	Th/K	1.28
Topography	slope	1.60	slope	1.60
	transformed aspect	1.01	transformed aspect	1.01
	CTI	1.80	CTI	1.80
	distance to drainage	1.41	distance to drainage	1.40
	distance to wetland	1.62	distance to wetland	1.44
Modification	land use	2.25	land use	2.19
Landsat imagery	FPC	3.91	FPC	3.85
	NDVI	2.74	NDVI	2.72
	bare ground	2.45	bare ground	2.45
Location	Easting	**12.5**		
	Northing	2.68		

References S2

Alin A (2010) Multicollinearity. WIREs Computational Statistics 2(3):370-374 https://doi.org/10.1002/wics.84

Citakoglu H (2017) Comparison of artificial intelligence techniques for prediction of soil temperatures in Turkey. Theoretical and Applied Climatology 130(1):545-556 https://doi.org/10.1007/s00704-016-1914-7

O'Brien RM (2007) A caution regarding rules of thumb for variance inflation factors. Quality & Quantity 41(5):673-690 https://doi.org/10.1007/s11135-006-9018-6

Vu DH, Muttaqi KM, Agalgaonkar AP (2015) A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables. Applied Energy 140:385-394 https://doi.org/10.1016/j.apenergy.2014.12.011

Supplementary Material S3 – Additional information on artificial neural network model architecture

Artificial neural network (ANN) models are used to find patterns in complex data through an iterative 'train and learn' process. Training is the process of recognising the associative patterns and relationships between the response data and environmental predictor surfaces; learning is the process of finding the least amount of error in the patterns and relationships. Training passes information forward through the network via weights and bias; learning informs the network by adjusting weights to reduce the largest errors.

Lek et al. (1996), Olden et al. (2008), Özesmi et al. (2006) and Zhang et al. (1998) provide examples of working with ANN models within an ecological context and Bishop (1995), Fielding (1999) and Haykin (2009) provide comprehensive and authoritative texts on the principles of neural networks.

Many types of ANN models exist (see Bishop 1995; Ripley 1996), the most frequent architecture used in ecological applications is a multi-layer perceptron (MLP) feed-forward network with one hidden layer trained by either back-propagation (Rumelhart et al. 1986), or as in our application, the Broyden-Fletcher-Goldfarb-Shanno BFGS algorithm (see Bishop 1995). The MLP framework is used to construct ANN models because their architecture is robust when handling large volumes of complex data.

Below, we discuss the important components of the ANN architecture – observation data, the training algorithm, activation functions and weight decay, predictive output and sensitivity analysis.

Observation Data

Site locations (X and Y coordinates) were assembled with the intersected point values for each of the environmental and disturbance predictor surfaces (software specifications are outlined in Supplementary Material S4). This matrix formed the response data used to train the ANN. It is generally advised that the dataset contains ten times more observations than predictor surfaces

Outlying Observations

Even though ANN models can effectively discount outliers, isolated data points undermine the split-sample hold-out strategy because there are too few cases allocated to either the train, test or hold-out subsets. This, in turn, has a large effect on the error (Bishop 1995) and consequently, poor predictions can be expected when the verification data contain values outside of the range of those used for training or testing (Maier and Dandy 2000). We visually inspected for outlier values. Heuristically identifying outliers is an important step in the three-way train:test:hold-out modelling framework. Even though ANN models can effectively discount outliers (Bishop 1995), outliers undermine the split-sample hold-out strategy because there are too few cases allocated to either the train, test or hold-out subsets. This, in turn, has a large effect on the error (Bishop 1995) and consequently, poorer predictions can be expected when the hold-out data contain values outside of the range of those used for training or testing (Maier and Dandy 2000). Summed cover outliers (most likely due to human error in making visual estimates of cover) were identified and excluded from cover models.

Training algorithm

Training the neural network involves an error algorithm, which finds a set of connection weights that produces an output signal with a small error relative to the observed output. There are many texts that explain the differences between gradient descent and quasi-Newtonian methods, but the fundamental difference is that back-propagation is a gradient descent method and BFGS is a second derivative of the Newtonian method and uses a Hessian error matrix. In terms of applying ANN to ecological modelling, the difference is realised in computational penalties. BFGS require less computational power, which in turn delivers faster models.

Activation functions

Activation functions introduce nonlinearity into the model. It is the nonlinearity (that is, the capability to represent nonlinear functions) that make ANN models so effective. For both the hidden and output layers we used the negative exponential activation function because the output layers are continuous data, bounded by 0 at the lower with the potential for infinity at the upper end (Bishop 1995).

Weight decay

To avoid over-fitting the model during training, we used weight decay as an effective form of complex regularisation (Haykin 1994). Minimum weight decay was set to 0.0001 and the maximum to 0.001, which is regarded as weak regularisation. Weight decay trains the network's sum of squares error function by penalising larger weights (Statsoft Inc. 2013).

The input and target variables are scaled using linear transformations such that the original minimum and maximum of every variable is mapped to the range (0,1). This is an automatic preprocessing step within the Statistica program (Statistica Electronic Textbook 2020).

Predictions

We explored the Artificial Network Search (ANS) in the Statistica Artificial Neural Network (SANN) module within Statistica v10 software (Statsoft Inc. 2013). To curtail the influence of spatial autocorrelation we also explored multiple different partitions of the response data train, test and hold-out while keeping the architecture fixed. For each vegetation attribute, we retained the 25 highest performing models (assessed by the coefficient of determination R^2) and combined these models into an ensemble model referred to as “ensemble of random sub-samples”. The models used to build the ensemble of random samples are stored as Predictive Model Markup Language (PMML) files. PMML models are a standard format used to represent predictive models across many modelling platforms applications.

Sensitivity analysis

An indication of the importance of each predictor surface was illuminated using global sensitivity analyses. Sensitivity analyses were calculated after the model was trained and were assessed by comparing the change in the original model error compared to the error obtained by substituting a variable with ‘missing value’. The ratio of ‘original error’ to ‘error when the variable is substituted’, gives an indication of how sensitive the network is to each predictor surface (Statsoft Inc. 2013). Predictor surfaces with sensitivity values less than one cause the model performance to degrade.

S3 Supplementary References

Bishop, CM. (1995). Neural Networks for Pattern Recognition. Oxford University Press.

Fielding, A. (1999). *Machine learning methods for ecological applications*. Springer Science & Business Media.

Haykin, SS. (1994). Neural Networks: A Comprehensive Foundation. 2nd edition. Prentice Hall PTR, New Jersey.

Lek S, Delacoste M, Baran P, Dimopoulos I, Lauga J, Aulagnier S (1996) Application of neural networks to modelling nonlinear relationships in ecology. Ecological Modelling 90(1):39-52 https://doi.org/10.1016/0304-3800(95)00142-5

Maier HR, Dandy GC (2000) Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling & Software 15(1):101-124 https://doi.org/10.1016/S1364-8152(99)00007-9

Olden J, Lawler Joshua J, LeRoy Poff N (2008) Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83(2):171-193 https://doi.org/10.1086/587826

Özesmi SL, Tan CO, Özesmi U (2006) Methodological issues in building, training, and testing artificial neural networks in ecological applications. Ecological Modelling 195(1–2):83-93 https://doi.org/10.1016/j.ecolmodel.2005.11.012

Ripley, B. D. 1994. Neural networks: a review from a statistical perspective: comment. Statistical Science 9:45-48.

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533-536 https://doi.org/10.1038/323533a0

Statsoft Inc. 2013. Electronic Statistics Textbook. http://www.statsoft.com/textbook, Tulsa, Oklahoma.

Zhang G, Patuwo BE, Y Hu M (1998) Forecasting with artificial neural networks: the state of the art. International Journal of Forecasting 14(1):35-62 https://doi.org/10.1016/S0169-2070(97)00044-7

Supplementary Material S4 – Software to create data input tables, prediction output tables and manage multiple tables

Creating the training matrix and the prediction matrix and interpreting the prediction output into a raster map was accomplished with the following suite of software tools. There are 32-bit and 64-bit versions of these tools. These are all developed under Microsoft.NET and require Microsoft.NET application Framework version 4.0.

- NNExtractFromBinaries.exe (see manuscript Figure 2 Step 2.2 and Step 4.2)
- NNToolsExtractor.exe (see manuscript Figure 2 Step 4.1)
- NNToolsPrediction.exe (see manuscript Figure 2 Step 5.1)

NNExtractFromBinaries.exe

To build the training matrix, for each of the response data (points), the underlying values from each of the predictor (raster) surfaces were extracted using the NNExtractFromBinaries.exe application (Figure 2 Step 2.2). A comma-delimited text file, formatted as Site_ID, X_Coordinate, and Y_Coordinate was used as input into the NNExtractFromBinaries.exe application. The output table was also a comma-delimited table with the first three columns formatted as the input table and the comma-delimited values for each of the predictor surfaces in subsequent columns. The same tool was used to build the prediction matrix (Figure 2 Step 4.2), using the cell centroids from each of the 100 m grid cells in the study area and the same suite of predictor surfaces.

The predictor surfaces must all be in ESRI Binary Export Format (.flt/.hdr) and can be loaded into the listView in the lower half of the application. The csv table is supplied as the input file to train the ANN models.

NNToolsExtractor.exe

The extent of the study area was prepared as a 100 m raster grid in ESRI Binary Export Format. The centroid of each grid cell in the spatial extent becomes a new site for which the trained network makes a prediction (Figure 2 Step 4.1). These cell centroids are extracted to create the first section of the prediction matrix. The NNToolsExtractor.exe application created a table with the following Format _ID, X_Coordinate, Y_Coordinate where the X and Y coordinates of all the centroids for each of the 11.5 million cells in the study area. Once the X Y locations for every new grid cell are nominated, this table can then be submitted to the NExtractFromBinaries.exe application to extract the predictor data for each record (Figure 2 Step 4.2).

NNToolsPrediction.exe

Once the predictions have been acquired via the PMML output from the model, the resulting table can be used as input to the NNToolsPrediction.exe application (Figure 2 Step 5.1). The user can select the X and Y fields and importantly, the prediction field and by using the mask grid as a domain will create an ESRI Binary Export grid that will have each valid data cell coded with the predicted value from the ANN model.

Supplementary Material S5 – Equations for calculating root mean squared deviation (RMSD) and mean absolute error (MAE)

The RMSD estimate represents the mean deviation of predicted foliage cover or species richness values with respect to the observed foliage cover or species richness values. In addition, RMSD is most useful for evaluating models as it represents the absolute measure of fit to the 1:1 line and reports the prediction error in the same units as the data (i.e. summed cover or richness) (Piñeiro et al. 2008).

$$RMSD = \sqrt{\frac{1}{n-1}\sum_{i-1}^{n}(\hat{y}_i - y_i)}^{\;2} \qquad \text{(Equation 1)}$$

Where, $\hat{y}_i$ are the predicted cover values

y_i are the observed cover values and

n is the number of observations.

$$MAE = \left[n^{-1}\sum_{i=1}^{n}|e_i|\right] \qquad \text{(Equation 2)}$$

Mean absolute error is the average absolute vertical or horizontal distance between each point in a scatter plot and the 1:1 line. MAE is a natural measure of the average error, which can sometimes be simpler for end-users to interpret. Furthermore, each error contributes to MAE in proportion to the absolute value of the error (Willmott & Matsuura 2005).

S5 Supplementary References

Piñeiro G, Perelman S, Guerschman JP, Paruelo JM (2008) How to evaluate models: observed vs. predicted or predicted vs. observed? Ecological Modelling 216(3–4):316-322 https://doi.org/10.1016/j.ecolmodel.2008.05.006

Willmott CJ, Matsuura K (2005) Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Research 30(1):79-82 https://doi.org/10.3354/cr030079

Supplementary Material S6 – Sensitivity analysis showing the sensitivity averaged over 25 ensembles of ANN models for ten vegetation attributes

Table S6: Sensitivity analysis showing the sensitivity averaged over 25 ensembles of ANN models for ten vegetation attributes. Only the highest three sensitivity values for each vegetation attribute are shown. Sensitivity values are unitless.

Vegetation Attribute	Land use	Great Soil Group	Foliage Projective Cover	Clay %	Precipitation/ Evaporation	Isothermality	Gravity	Distance to drainage
Tree cover	2.15	1.04	1.32					
Shrub cover	1.23	1.03	1.02					
Grass cover	1.38	1.05			1.01			
Forb cover	1.07	1.0	1.0					
Total cover	3.29	1.04	1.12					
Tree richness	2.02	1.08		1.03				
Shrub richness	1.84	1.12	1.03					1.02
Grass richness	2.02	1.05				1.02		
Forb richness	1.98	1.08					1.02	
Total richness	3.49	1.1		1.01				

Supplementary Material S7 – Results to assess for spatial autocorrelation - Moran's Index

Failing to account for spatial autocorrelation can violate the assumption of independence or lead to errors in model interpretation. To evaluate if our approach to managing spatial autocorrelation (e.g., specifying the minimum distance between sites, removing collinear predictor variables and building ensembles of random samples) was an effective strategy, we calculated Moran's Index for the model residuals. Table S7 shows the results.

Table S7: Results of the local analysis on standardised residuals to assess spatial autocorrelation.

Vegetation Attribute	Moran's Index	z-score	p-value	Variance
Tree cover	0.001	0.001	1.00	0.965
Shrub cover	0.010	0.011	0.99	0.964
Grass cover	0.013	0.013	0.99	0.964
Forb cover	0.082	0.084	0.93	0.956
Total Cover	0.008	0.009	0.99	0.965
Tree richness	0.570	0.704	0.48	0.654
Shrub richness	0.231	0.286	0.77	0.654
Grass richness	0.338	0.418	0.68	0.655
Forb richness	-0.205	-0.253	0.80	0.655
Total Richness	0.597	0.738	0.46	0.655

Moran's Index of -1 indicates model residuals are significantly dispersed. Moran's Index of 1 indicates model residuals are clustered. Where values are close to zero, the pattern is considered random. Z-scores are standard deviations. If, for example, a z-score = 2.58, the result is 2.58 standard deviations. Both z-scores and p-values are associated with the standard normal distribution (Figure 1 in S7). Very high or very low Moran's Index, z-scores or very small p-values are found in the tails of the normal distribution and indicate that model residuals are either dispersed or clustered.

- Z-scores greater than 2.58 or less than -2.58 show there is less than 1% likelihood that either the clustered or dispersed patterns are a result of random chance.
- Z-scores between 1.96 to 2.58 and -1.96 to -2.58 show there is less than 5% likelihood that either the clustered or dispersed patterns are a result of random chance.

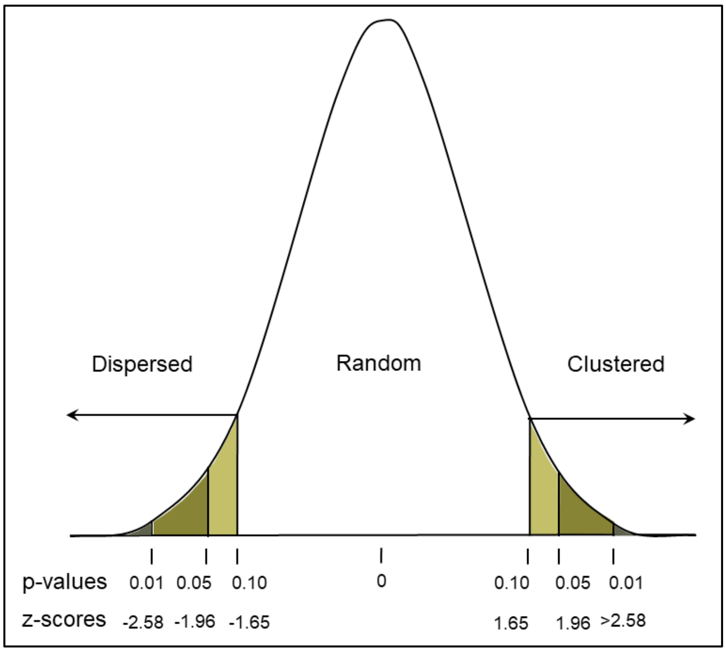

Figure S7 – Schematic diagram illustrating ArcGIS Spatial Statistics toolbox (ArcGIS v10.4). tool calculates the critical value z-score and significance level p-value to indicate if model residuals are dispersed, random or clustered

Supplementary Material S8 – The spatially explicit maps and accompanying residual maps for each vegetation attribute

Figures S1-S5 Maps of predicted vegetation attributes and their standardised residual error. The standardised residual error between predicted and observed values was calculated for each prediction at the site and interpolated across the landscape using ordinary kriging in ArcGIS Spatial Analyst. These values represent areas where the model has under-estimated (positive values) or over-estimated (negative values). Residual error maps were classified so that residual error (-1 and 1) are white; increasing residual error (between 1 to 2 and -1 to -2) as shaded lighter, and areas where there is higher uncertainty (less than -3 or greater than 3) are darker.

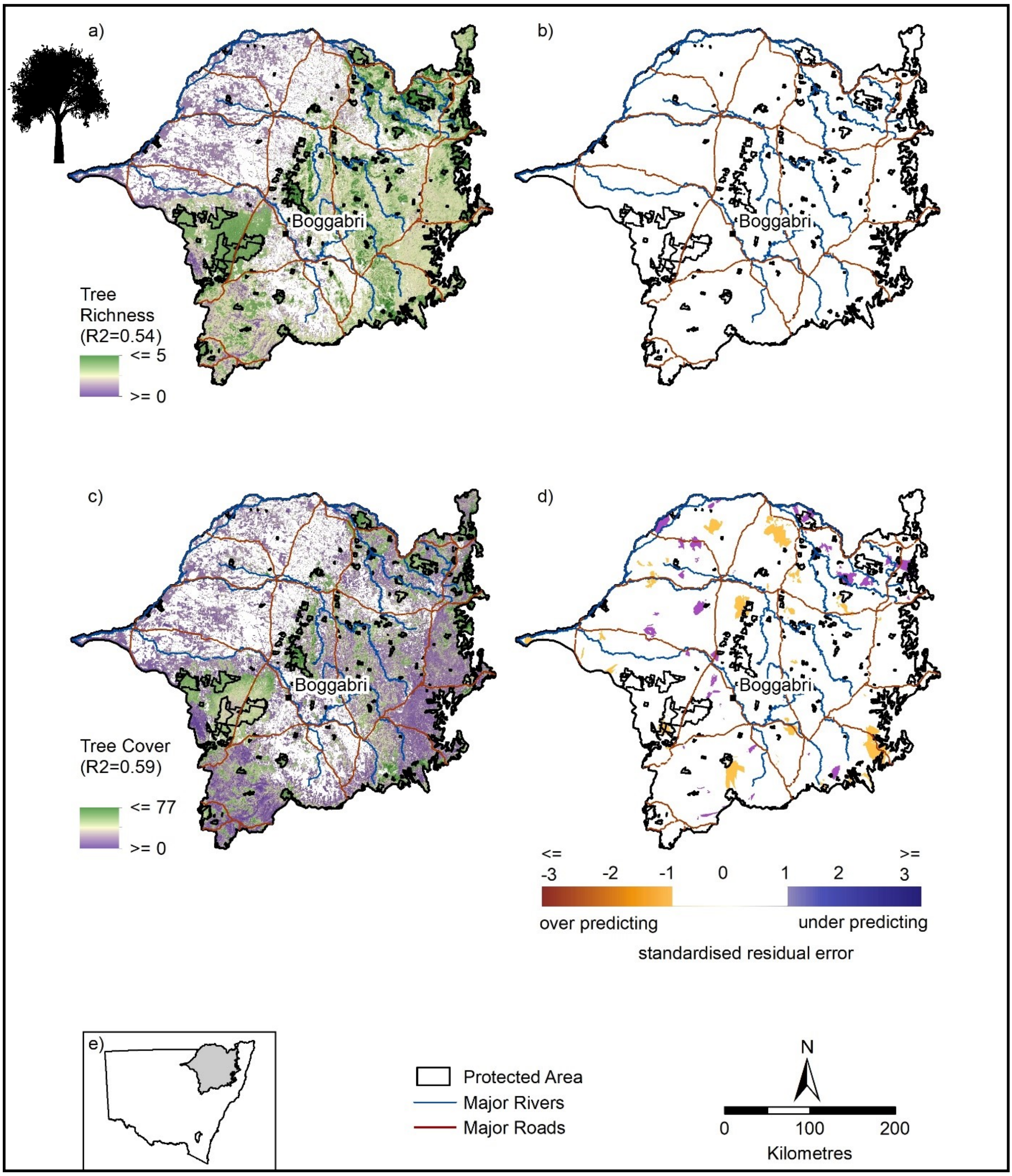

Figure S1 a) native tree richness; b) standardised residual error for native tree richness; c) native tree cover (%); d) standardised residual error for native tree cover; and e) inset map showing study area

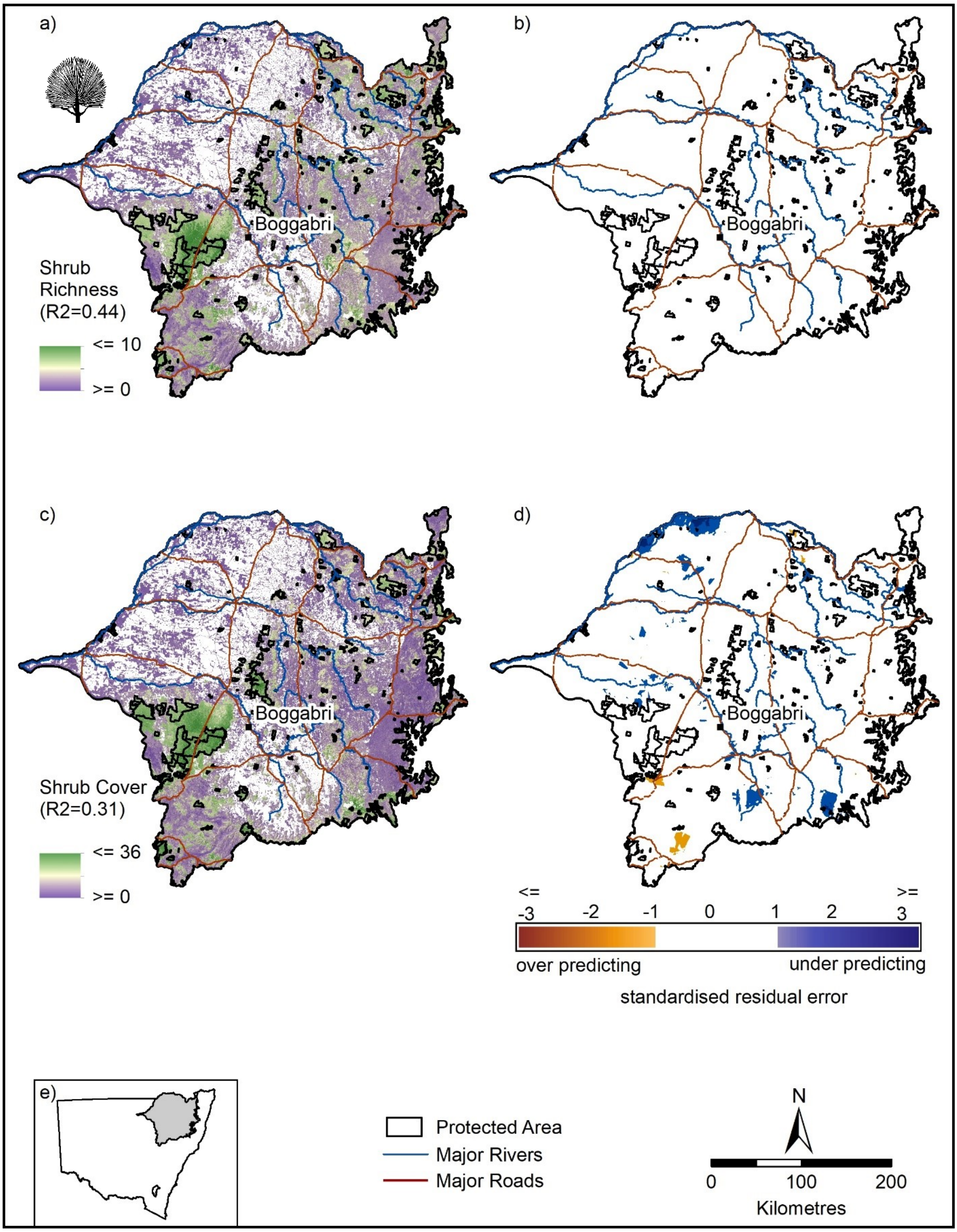

Figure S2 a) native shrub richness; b) standardised residual error for native shrub richness; c) native shrub cover (%); d) standardised residual error for native shrub cover and e) inset map showing study area

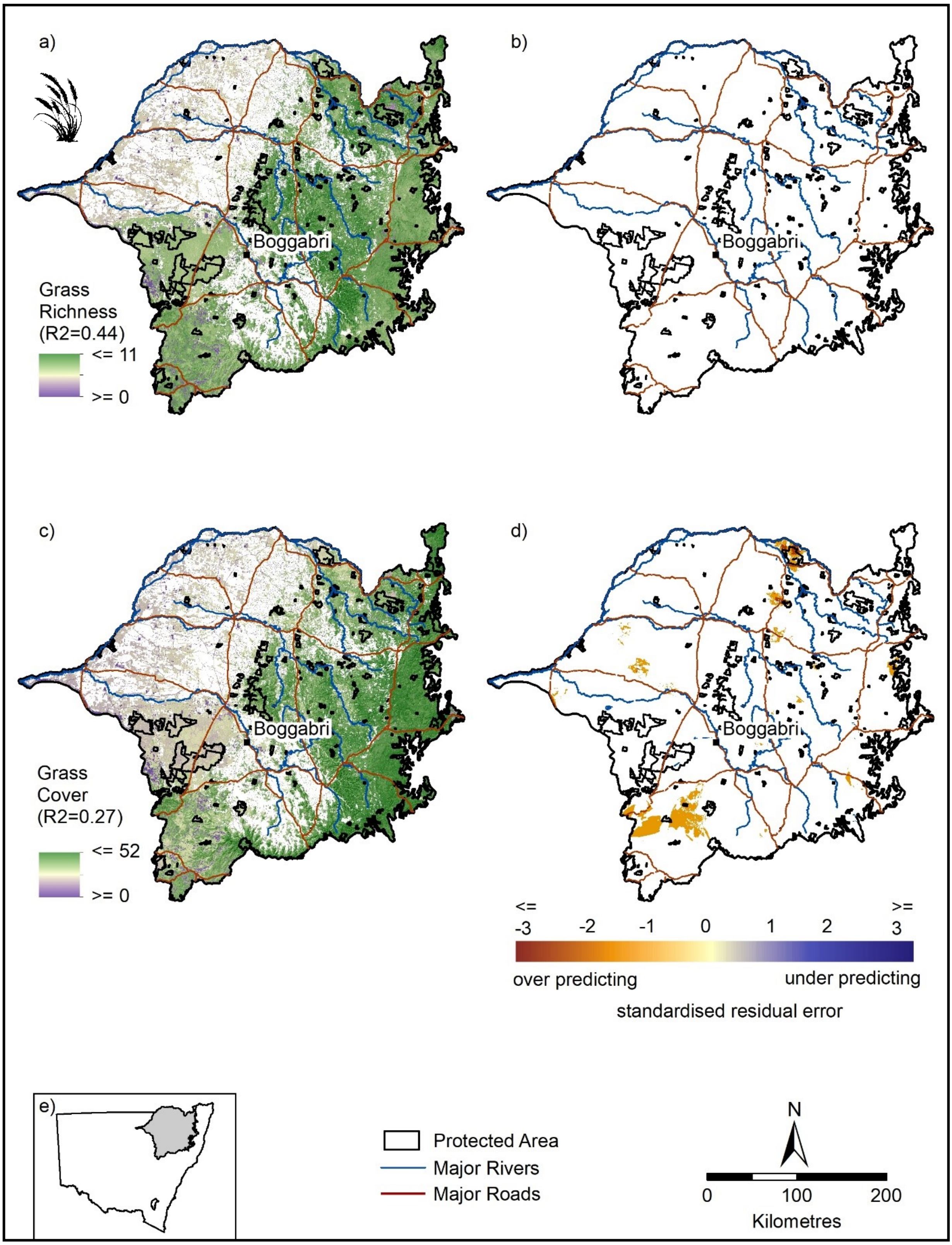

Figure S3 a) native grass richness; b) standardised residual error for native grass richness; c) native grass cover (%); d) standardised residual error for native grass cover and e) inset map showing study area

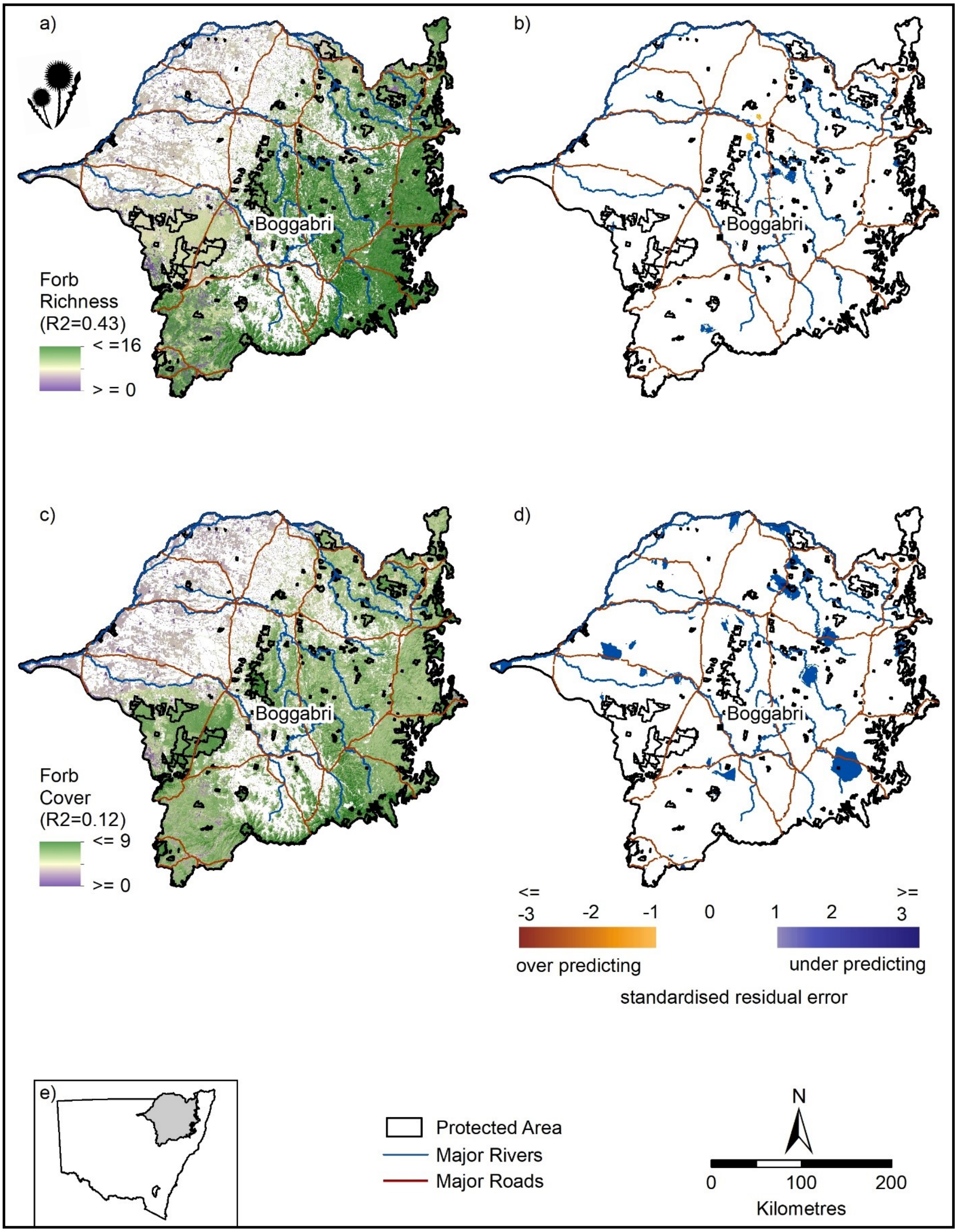

Figure S4 a) native forb richness; b) standardised residual error for native forb richness; c) native forb cover (%); d) standardised residual error for native forb cover and e) inset map showing study area

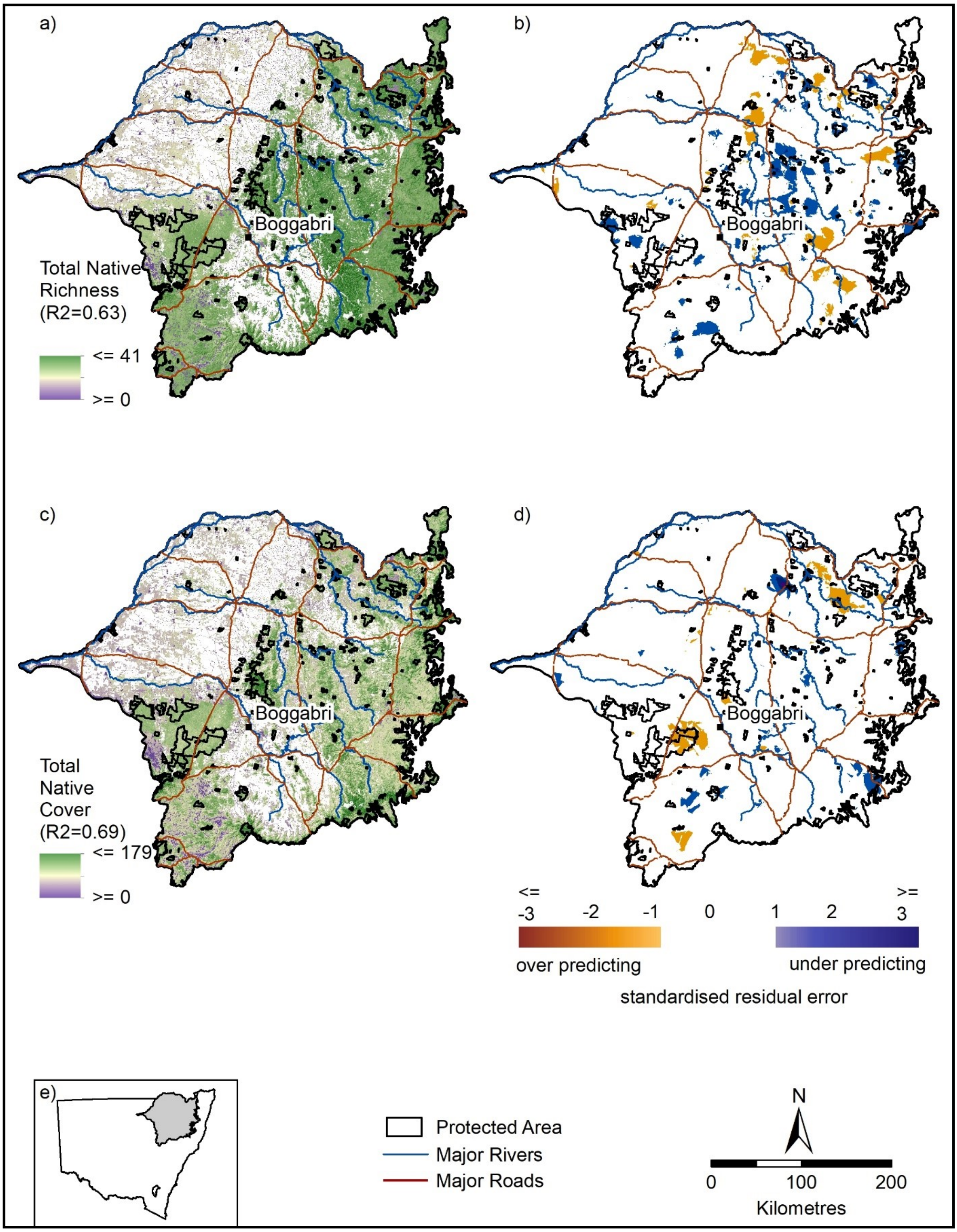

Figure S5 a) total native richness; b) standardised residual error for total native richness; c) total native cover (%); d) standardised residual error for total native cover and e) inset map showing study area