Using community-level modelling to map levels of biodiversity significance in the Pilbara Bioregion

Kristen J Williams[1], Thomas Harwood[1], Justin Perry[1], Genevieve Perkins[1] Yang Liu[2] and Simon Ferrier[1]
1. CSIRO Ecosystem Sciences, Canberra; 2. CSIRO Computational Informatics, Canberra

Final Report, 30th November 2013

Prepared for BHP Billiton Iron Ore

Copyright and disclaimer

Important disclaimer

Document management

Version	Date	Distribution
Draft Final Report	21st July 2013	CSIRO project team (SF, KJW, TH, JP, GP)
		BHPBIO via Mark Vile
Revised draft Final Report	21st October 2013	CSIRO project team (SF, KJW, TH, JP, GP, YL)
		CSIRO internal review – Justine Murray, Dutton Park, Brisbane
		BHPBIO via Mark Vile
Final Report	30th November 2013	CSIRO project team (SF, KJW, TH, JP, GP, YL)
		BHPBIO via Mark Vile

Suggested citation

Williams KJ, Harwood T, Perry P, Perkins G, Liu Y, Ferrier S (2013) *Using community-level modelling to map levels of biodiversity significance in the Pilbara Bioregion*. Unpublished Report for BHP Billiton Iron Ore. CSIRO Ecosystem Sciences, Canberra.

Cover page

"Remembering the Pilbara", a painting by Sharon Jack, http://www.sharonjack.com.au/; used with permission.

Foreword

BHP Billiton Iron Ore (BHPBIO) is proposing a regional approach to its environmental impact assessment and management for mining iron ore in the Pilbara region, Western Australia. This approach is being undertaken over a relatively large area across which modelling tools can assist the prediction of potential impacts and conservation outcomes. An early identification of environmental significance will usefully inform planning and management.

CSIRO was engaged to undertake an assessment of spatial patterns in the distribution of biodiversity, and associated levels of biodiversity significance, across the Pilbara Bioregion. This work employs state-of-the-art techniques for community-level modelling and biodiversity assessment, integrating best-available existing biological and environmental data for the region.

The outputs of this assessment are maps of "biodiversity significance" across the Pilbara bioregion, representing the potential for a given location to harbour a concentration of species narrowly distributed beyond that location, due to natural patterns of endemism and/or anthropogenic habitat degradation.

These outputs can be used to assess potential effects of disturbance and management on biodiversity from a whole-of-bioregion perspective.

<name>

BHP Billiton Iron Ore

Contents

Figures

Tables

Acknowledgments

Author contributions

SF conceived the research; JP, KJW and SF coordinated activities. GP and KJW compiled and vetted the biological data; KJW and TH compiled the environmental data; KJW and SF developed the GDM models; KJW and YL developed the GAM models; JP developed the condition models with input from KJW and SF on the interim condition model design. SF and TH developed the biodiversity significance models; TH developed the biodiversity significance mapping. SF, KJW and JP wrote the draft report. KJW and SF wrote the final report and all authors contributed equally to the revisions.

Executive Summary

Purpose

BHP Billiton Iron Ore (BHPBIO) engaged CSIRO to undertake an assessment of spatial patterns in the distribution of biodiversity, and associated levels of biodiversity significance, across the Pilbara Bioregion. The "overall biodiversity" perspective adopted in this study, through the use of community-level data and modelling techniques, is purposely intended to complement other recent work commissioned by BHPBIO focusing on modelling and assessment of individual species of particular conservation concern.

Overall approach

This work employed state-of-the-art techniques for community-level modelling and biodiversity assessment, integrating best-available existing biological and environmental data for the region, as outlined below.
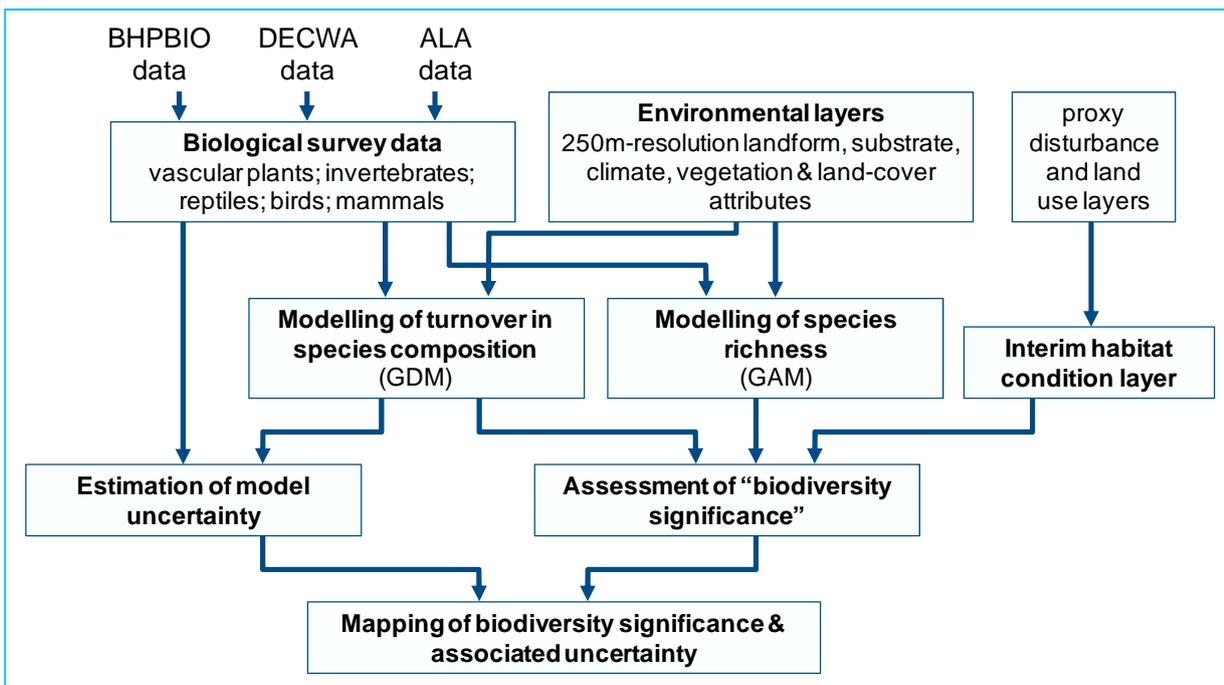
Above: overview of project activities.

Biological survey data provided by BHPBIO (concentrated mostly in or around BHPBIO tenements) were combined with data from a systematic and representative survey of over 300 sites across the Pilbara Bioregion conducted by the Western Australian Department of Parks and Wildlife (DPaW), and with species locality records accessed from the Atlas of Living Australia (ALA). These combined data were subjected to rigorous vetting procedures (see Section 2) before being employed in subsequent modelling analyses of species compositional dissimilarity using generalised dissimilarity modelling (GDM) and richness using generalised additive modelling (GAM) (Section 3) that underpinned the assessment of biodiversity significance (Section 5).

Potential data sources, and techniques, for mapping habitat condition across the Pilbara were investigated and trialled, including an approach using remote sensing. The interim condition layer eventually employed in this study was estimated from a combination of best-available proxy variables relating to various

pressures (grazing, mining, residential and transport infrastructure) and ameliorating factors (protected areas and steep slopes).

Finally, the modelled patterns of species richness and compositional dissimilarity were brought together with the interim condition layer to estimate "biodiversity significance" across the bioregion, in terms of the potential for a given location to harbour a concentration of species narrowly distributed beyond that location, due to natural patterns of endemism and/or anthropogenic habitat degradation.

Datasets

The digital data resulting from these analyses (9-second grids) have been made available to BHPBIO via CSIRO's Data Access Portal (www.data.csiro.au). These comprise predicted species compositional dissimilarity (Williams *et al.* 2013a) and richness (Williams *et al.* 2013b) for each biological group, habitat/landscape condition (Perry *et al.* 2013b, a) and biodiversity significance for each biological group and overall groups (Harwood *et al.* 2013).

Limitations

The data requirements for community-level modelling and the need to merge data between different sources determined which data could be used. The merging of data required that all taxon names be related by a common nomenclature and at a common taxonomic rank (e.g., species). The BHPBIO data of least use were the invertebrate group which were commonly identified only to genus level. We therefore relied upon the invertebrate data from DPaW Pilbara biological surveys (George *et al.* 2011) for this analysis.

The data requirements for compositional turnover and richness modelling are different. Although comprehensive and representative survey data are preferred, such data are typically limited in extent. While generalised dissimilarity modelling of compositional turnover is relatively robust to variation in data quality where there is also good geographic coverage, the generalised additive modelling of species richness requires much greater rigour at the site-level of the survey data. For the richness modelling, we used only those records from the BHPBIO data for which a specified sampling method was provided as an indicator of site-level sampling quality. This substantially reduced the spatial extent of data available for this analysis and, with a mix of survey type without suitable covariates, limited model effectiveness.

The vascular plant data underpinning the community-level models were largely drawn from BHPBIO sources, supplemented by the Atlas of Living Australia. The systematic surveys of vascular plants (George *et al.* 2011) were unavailable because species identifications were incomplete. The spatial bias resulted in high confidence in model outputs surrounding BHPBIO's mining tenements, but less confidence in predictions in distinct environments with limited survey coverage. Nonetheless, quite distinct assemblages are apparent across the Pilbara between the western coastal zone, north and inland that broadly align with the subregions, demonstrating the robustness of the GDM approach, in particular.

Limitations in the predictive capacity of the community-level models were expressed in terms of scaled uncertainty represented by varying levels of transparency as a white overlay that is always associated with the derived biodiversity significance map outputs. Two quite different types of uncertainty are presented for the two community-level models. For compositional dissimilarity, the uncertainty represents the density of survey coverage in GDM-scaled environmental space. Areas of high sampling coverage generally correspond with higher confidence in GDM-modelled environmental space, and are also areas where additional surveys would improve model reliability. For species richness, the uncertainty is a statistically-based estimate of the standard error of the predicted value from the GAM model. When both community-level models are used in the calculation of biodiversity significance, the two estimates of uncertainty are used together to spatially represent the limitations of the output for decision making.

Remote sensing has great potential for assessing habitat condition but a greater effort than available in this project is required to achieve a credible output for use in biodiversity assessment. Contrasting with a

traditional remote sensing definition of condition, which considers change between images over short periods of time, our definition of habitat condition aims to take into account the intrinsic capacity for biodiversity to persist dynamically over longer time scales in the presence of natural disturbances. The quantification and separation of natural and anthropogenic drivers of disturbance are therefore critical to this assessment. In lieu of such an application for the Pilbara, we developed an interim estimation of habitat condition using proxy variables, which provides a good indication of regional differences in habitat condition between sites but requires validation and integration with remote sensing for use locally.

Considering the limitations in the modelling of species richness and the local application of the interim estimation of habitat condition, we present six different formulations of biodiversity significance – with and without species richness included, and with and without habitat condition included. Ideally, biodiversity significance would be calculated with both species richness and habitat condition included, indicating the relative importance of a site for the persistence or retention of biodiversity due to its local uniqueness or because other examples have been removed or are degraded in condition. These outputs were derived to demonstrate both how biodiversity significance is influenced by the incorporation of richness and habitat condition, and also to provide options for critical decision making where reliability of the underpinning data is paramount.

Key findings and recommendations

1. The data compiled by BHP for environmental assessment is a rich source of information, albeit often restricted to tenement areas. We focussed on certain aspects of these data for the analyses presented here; specifically, spatially referenced observations with taxonomic identifications at least to the species level that could be assigned an accepted name based on National censuses of plant and animal species. Two data sets presented particular limitations. Firstly, the invertebrate collections were mainly identified at the generic level or from poorly known groups based on targeted sampling, preventing their merger with the comprehensive surveys of beetles, ants and spiders across the Pilbara by the WA Department of Parks and Wildlife. Secondly, the comprehensive surveys of vascular plants across the Pilbara by the WA Department of Parks and Wildlife, which would help moderate the sampling bias associated with the surveys conducted by BHPBIO and their consultants, were unavailable at the time of this assessment. Finally, a further limitation of the BHPBIO data was the inability to easily distinguish the comprehensively surveyed sites for a particular biological group as reference sites for biodiversity modelling (richness and compositional dissimilarity). Different observation sources can be merged so long as covariates describing their similarities and differences can be generated and include in the modelling, to essentially weight their importance. This requires more detail about each survey's methodology and purpose, generally contained in survey reports.

2. The data underpinning the biodiversity significance models presented here can be enhanced through a concerted effort to compile all available surveys from among development interest groups across the region. Rather than a once-off compilation, we suggest data provider agreements with the Atlas of Living Australia and/or the Terrestrial Ecosystem Research Network who have established the protocols for aggregation and public use, or other appropriate entity to efficiently mediate data federation. Researchers, industry and consultants alike can then access the data, add value or assess the most cost-efficient requirements for filling knowledge gaps (e.g., through further survey) for a given purpose.

3. Species richness is a critical determinant of Biodiversity Significance, although secondary to models of compositional dissimilarity. The limitations of the current GAM models can be addressed through a more thorough examination of the available data sources and consideration of different statistical approaches. However, there is presently no specific best practice statistical approach and before investing in further analyses, gaps in the underpinning data require improvement through a more thorough examination of available sources as suitable 'reference sites' for modelling, and a targeted program of comprehensive surveys to fill gaps in environmental and taxonomic coverage of key indicator groups for conservation assessment.

4. Reliable assessment and monitoring of biodiversity habitat condition is also a critical determinant of Biodiversity Significance; and especially for any assessment of mining impacts and offsets. We present a novel framework toward solving this problem (Donohue *et al.* 2013) and outline a process for trialling this in the Pilbara (page 51). The successful application of this modelling framework, along with a network of monitoring sites to support calibration, will help integrate *in situ* and remote sensing data enabling a more reliable assessment of habitat condition applied at both regional and local scales for continuous reporting on the status of biodiversity and its use in scenario analyses.

5. Given current uncertainties in the assessment of species richness and site level habitat condition (outlined above), we can only recommend the use of Biodiversity Significance based on Equation 1 (natural uniqueness - Figure 35) and that incorporating regional condition (Equation 2 - Figure 36) for the current regional assessment process.

6. We recommend further interaction regarding the use of these outputs in BHPBIO's current and future strategic assessment work, to better resolve precisely which output to use for which purpose, or whether other variants of the measures presented here might be more applicable. For example, the map legends were designed for comparison within a group (with or without richness models included in the calculation), a different legend may be needed if the maps (e.g., Figure 35 and Figure 36) are used in isolation to ensure the maximum amount of information is conveyed.

7. Limitations in the biodiversity assessments identified here can be addressed through a program of research and development linked with BHPBIO's future planning and assessment needs, such as outlined in Figure 42.

1 Introduction

1.1 Purpose of this study

BHP Billiton Iron Ore (BHPBIO) engaged CSIRO to undertake an assessment of spatial patterns in the distribution of biodiversity, and associated levels of biodiversity significance, across the Pilbara Bioregion. This work has employed state-of-the-art techniques for community-level modelling and biodiversity assessment, integrating best-available existing biological and environmental data for the region. The "overall biodiversity" perspective adopted in this study, through the use of community-level data and modelling techniques, is purposely intended to complement other recent work commissioned by BHPBIO focusing on modelling and assessment of individual species of particular conservation concern.

1.2 Overall approach

We used community-level modelling techniques (Ferrier & Guisan 2006) to map patterns of richness and spatial turnover in terrestrial species composition for a wide range of biological groups – vertebrates, invertebrates and vascular plants. These patterns, combined with best-available information on habitat condition, were then used to estimate and map relative levels of biodiversity significance across the region. An overview of project activities is presented in Figure 1.

Biological survey data provided by BHPBIO (concentrated mostly in or around BHPBIO tenements) were combined with data from a systematic and representative survey of over 300 sites across the Pilbara Bioregion conducted by the Western Australian Department of Parks and Wildlife (DPaW), and with species locality records accessed from the Atlas of Living Australia (ALA). These combined data were subjected to rigorous vetting procedures (see Section 2) before being employed in subsequent analyses (Section 3 and 5).

Modelling of spatial patterns in the distribution of terrestrial biodiversity was performed by linking the biological data to a comprehensive set of mapped environmental variables collated by CSIRO, as part of a previous Pilbara modelling project undertaken through a partnership with DPaW. Two types of community-level model were fitted to the biological and environmental data, thereby enabling prediction (extrapolation) of biodiversity patterns across the entire bioregion:

- Modelling of the dissimilarity (turnover) in composition of species between each pair of surveyed locations, as a function of environmental differences between these locations.
- Modelling of the richness (number) of species occurring at each surveyed location as a function of environmental attributes at that location.

Potential data sources, and techniques, for mapping habitat condition across the Pilbara were investigated and trialled, including recent advances in remote sensing. The interim condition layer eventually employed in this study was estimated from a combination of best-available proxy variables relating to various pressures (grazing, mining, residential and transport infrastructure) and ameliorating factors (protected areas and steep slopes).

Finally, the modelled patterns of richness and compositional turnover were brought together with the interim condition layer to estimate the "biodiversity significance" of each and every location (9-second, approximately 250m, grid cell) across the bioregion, in terms of the potential for a given location to harbour a concentration of species narrowly distributed beyond that location, due to natural patterns of endemism and/or anthropogenic habitat degradation.
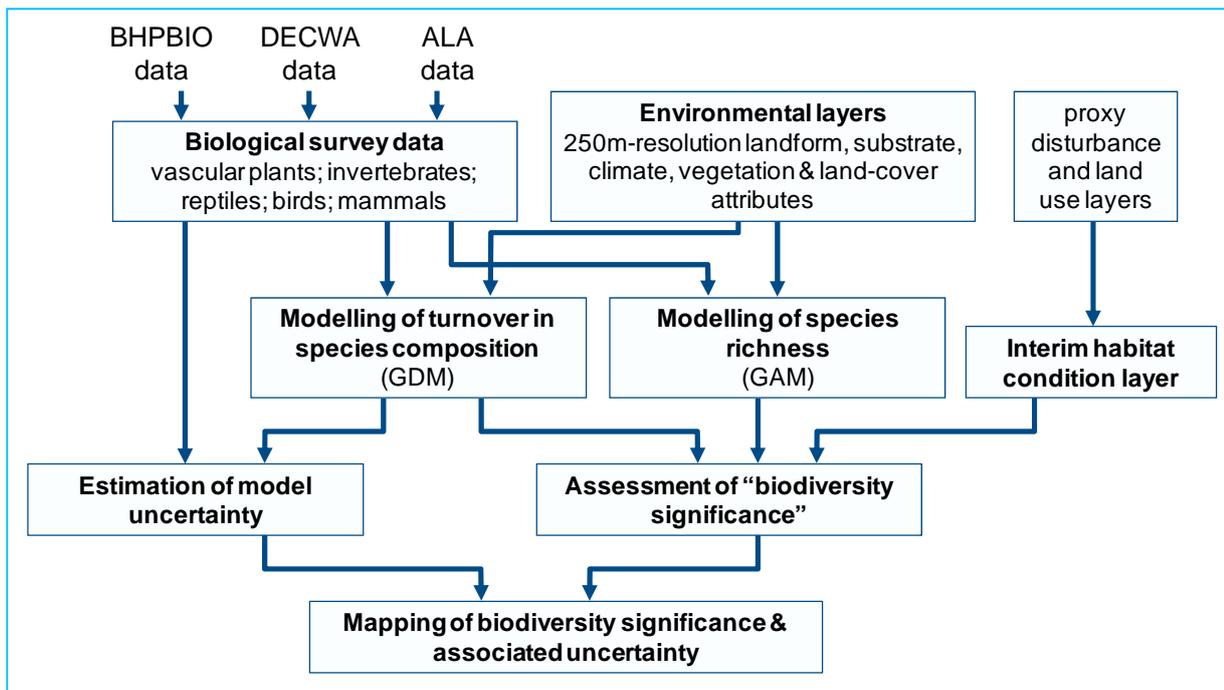
Figure 1. Overview of project activities.

1.3 Report outline

The remaining sections of this report follow the project activities summarised in Figure 1.

Section 2 "Data preparation" describes the compilation and filtering of biological and environmental data employed in this study.

Section 3 "Biodiversity model fitting" describes how the assembled biological and environmental data (from Section 2) were used to model spatial turnover in species composition, and spatial variation in species richness, across the bioregion.

Section 4 "Habitat condition mapping" describes the assessment of potential data sources, and techniques, for mapping habitat condition across the Pilbara, and the approach used to derive the interim condition layer employed in this study.

Section 5 "Biodiversity significance analysis" describes how the modelling of compositional turnover and richness (from Section 3) and mapping of habitat condition (from Section 4) were integrated to assess and map relative levels of biodiversity significance, and associated levels of uncertainty in this assessment, across the bioregion.

Section 6 "Discussion" provides an overall discussion of the results of this analysis, and the caveats pertaining to any interpretation of these results.

Appendices (A to F) provide supplementary information or additional details related to one or more of the above topics.

2 Data preparation

2.1 Introduction

Data aggregated from different sources for biodiversity modelling needs to be assessed as fit for purpose. Each record is matched to an accepted species name and location records filtered to reasonably match the spatial resolution of the proposed analysis. The geographic spread and density of data for each biological group is then assessed to determine whether there is sufficient information for analysis and at which level of the taxonomic hierarchy to apply the grouping. In this section, we outline the data sources, the approach taken to assess the suitability of the data for analysis, and present the resulting subsets of data used in subsequent analyses.

2.2 Study region

This study focuses on the Pilbara biogeographic region (DSEWPAC 2012) (Figure 2) which comprises four geomorphically distinctive subregions: undulating granite and basalt plains including significant areas of basaltic ranges (Chichester subregion); alluvial plains and river frontages (Fortescue Plains subregion); mountainous sedimentary ranges and plateaux dissected by gorges (Hamersley subregion); alluvial and older colluvial coastal and sub-coastal plains (Roebourne subregion) (Figure 3).
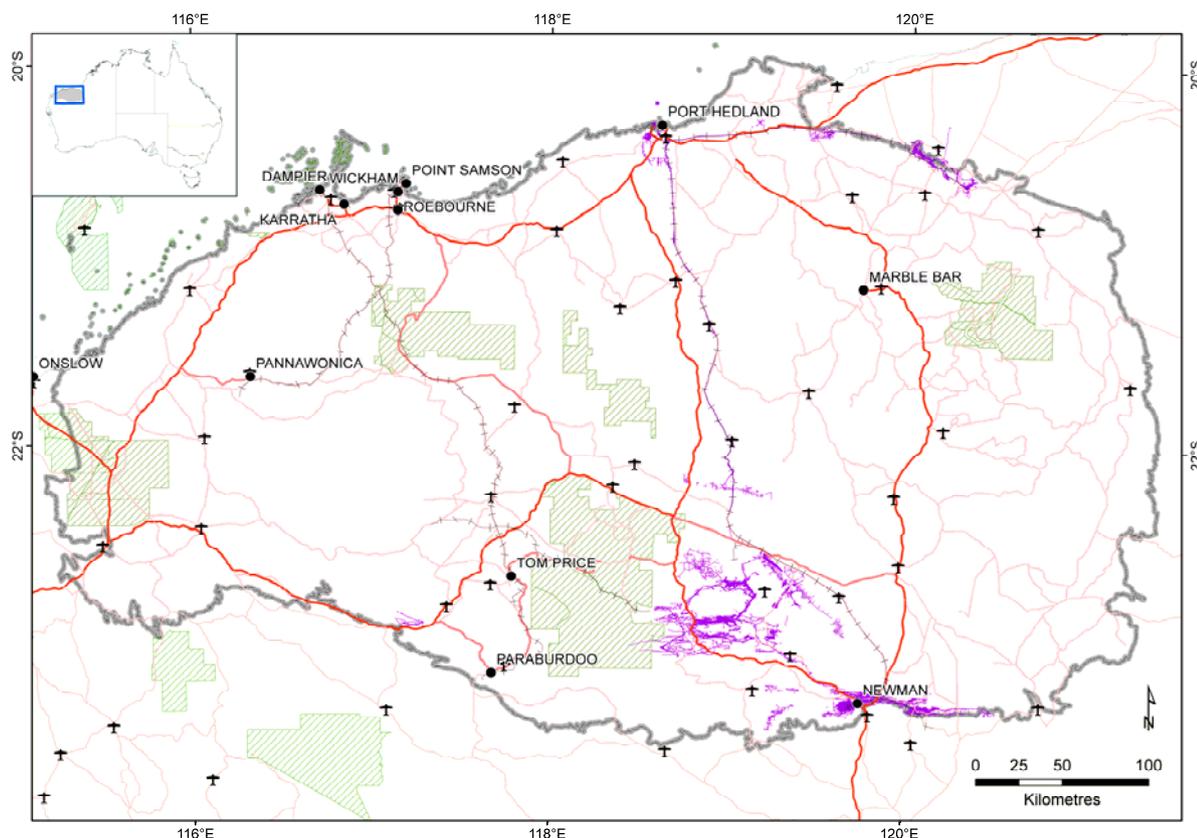
Figure 2. Pilbara study region showing the bioregion boundary, the BHPBIO infrastructure (roads, rail and current mining disturbance areas - violet), towns and other mapped infrastructure (road, rail, airfields), and protected areas in the National Reserve System (green hatched). The inset map of Australia shows the location of the Pilbara Bioregion.

The climate of the Pilbara region is semi-desert tropical, dominated by annual and inter-annual cycles of wetting and drying. January, February and March are the wettest months while September and October are the driest (McKenzie *et al.* 2009). There is considerable variation in rainfall between years due to cyclones that occasionally cross the coast. Rainfall intensity can be high with thunderstorms and cyclones generating high runoff volumes and fluvial patterns of erosion and deposition are apparent in the landform. The soil of the Pilbara region is generally skeletal having been derived *in situ* or deposited as colluviums or alluvium with colours reflecting the underlying parent material (McKenzie *et al.* 2009).

The analysis of biodiversity within the Pilbara bioregion is undertaken in the context of surrounding areas. Biological and environmental data are compiled for this extended area and models. This ensures that predictive model fitting process is not unduly influenced by truncated distributions at the margins of the study region.

Figure 3. Pilbara bioregion comprises four subregions: Roebourne, Chichester, Fortescue and Hamersley. The surrounding subregions are also shown.

2.3 Methods

2.3.1 BIOLOGICAL DATA SOURCES

The biological data were drawn from three sources, as outlined in Table 1. These data sources vary in their spread and density across the Pilbara biogeographic region. The BHPBIO data are focussed around the tenement areas of direct interest to mining and comprise both incidental sightings and systematic surveys (Figure 4). The DPaW data derive from a systematic and representative survey of about 300 terrestrial and 100 riparian and aquatic locations across the Pilbara biogeographic region (Figure 5).

A wide range of datasets were provided by BHPBIO and suitable datasets were compiled for analysis (Table 2). The data derived from the Atlas of Living Australia (ALA) aggregations are spread throughout the region (Figure 6). The ALA data were accessed via the biocache URL (http://biocache.ala.org.au/). A number of filters can be placed on the query to target the data of interest for download.

Table 1. Sources of biological records (taxa by location).

DATA SOURCE	BIOLOGICAL GROUP	NUMBER OF RECORDS (UNFILTERED)
BHP Billiton Iron Ore	Vertebrates	22,180
	Invertebrates	3840
	Vascular plants	96,959
Department of Parks and Wildlife, Government of Western Australia (DPaW)	Vertebrates	12,661
	Invertebrates	17,945
	Vascular plants	4275
Atlas of Living Australia partners (ALA)	Vertebrates	45,675
	Invertebrates	1122
	Vascular plants	13,456
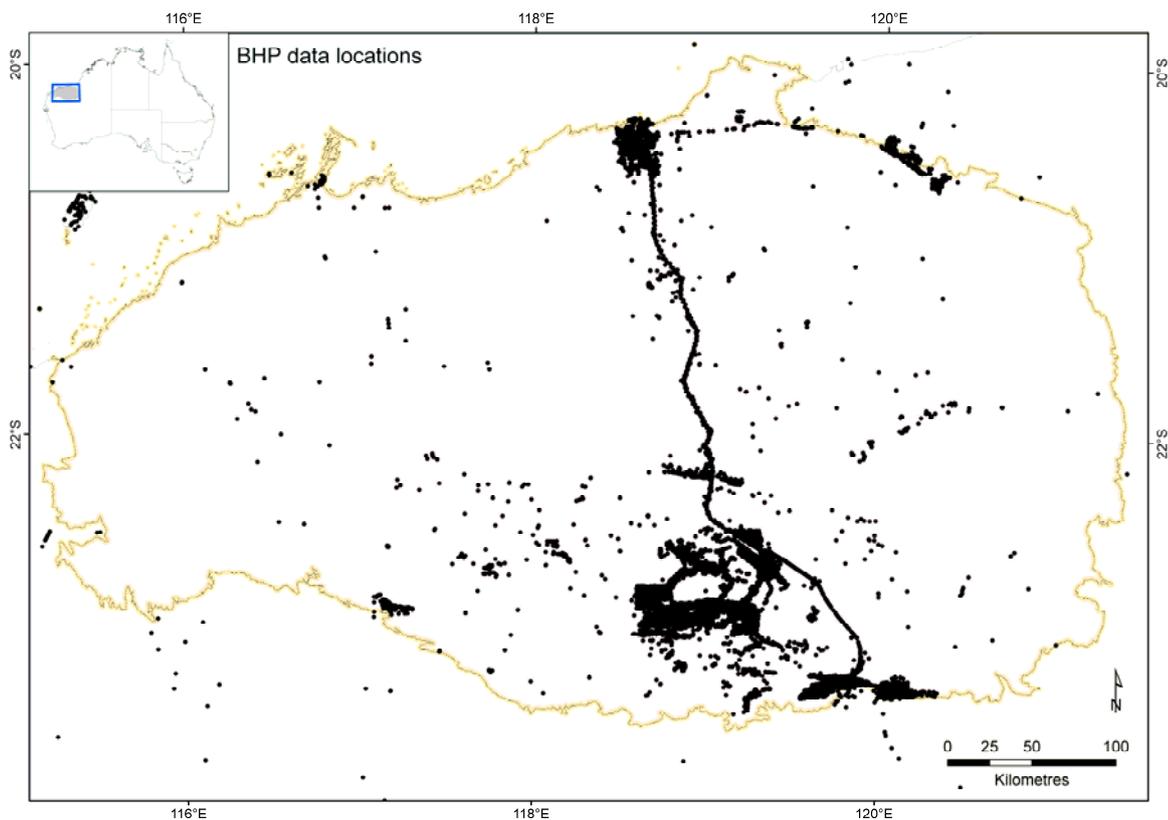
Figure 4. Locations of biological data sourced from BHPBIO.
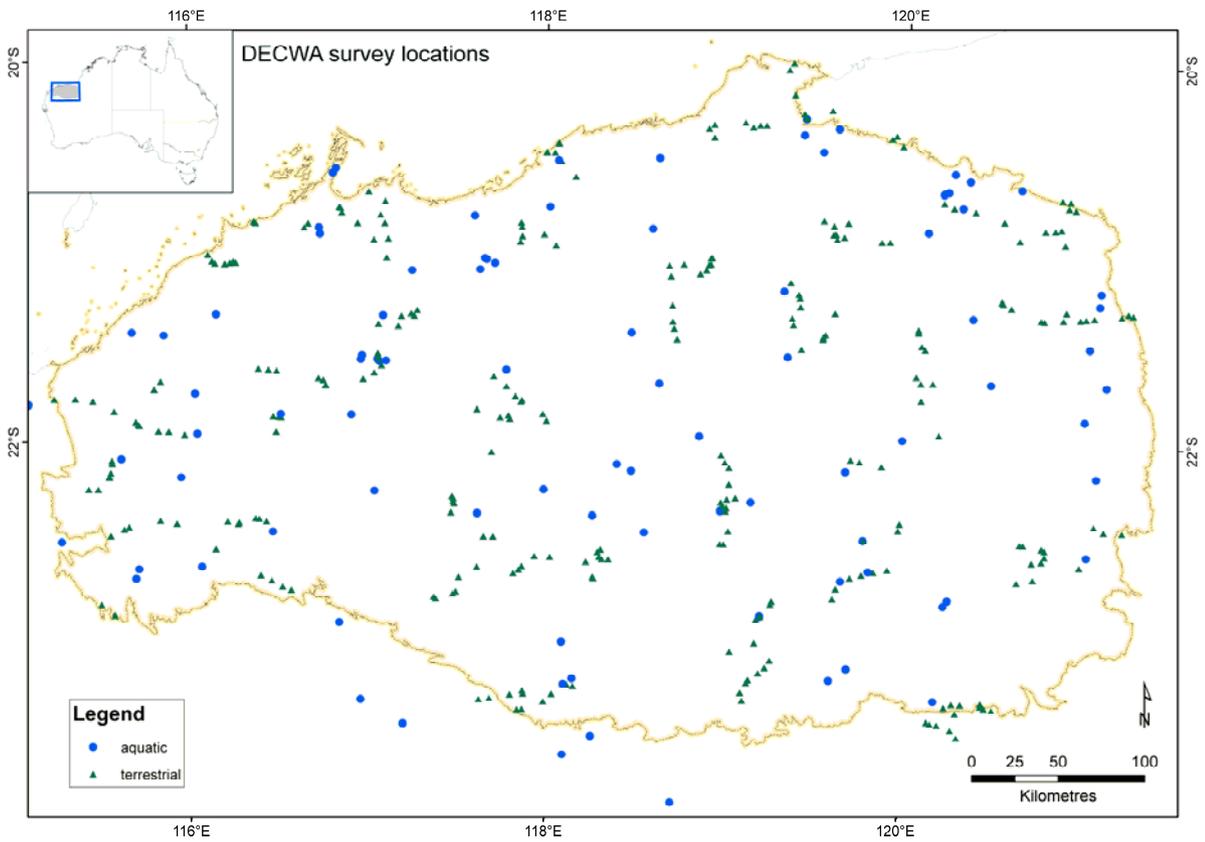
Figure 5. Locations of aquatic and terrestrial biological data sourced from DPaW (McKenzie *et al.* 2009; George *et al.* 2011).
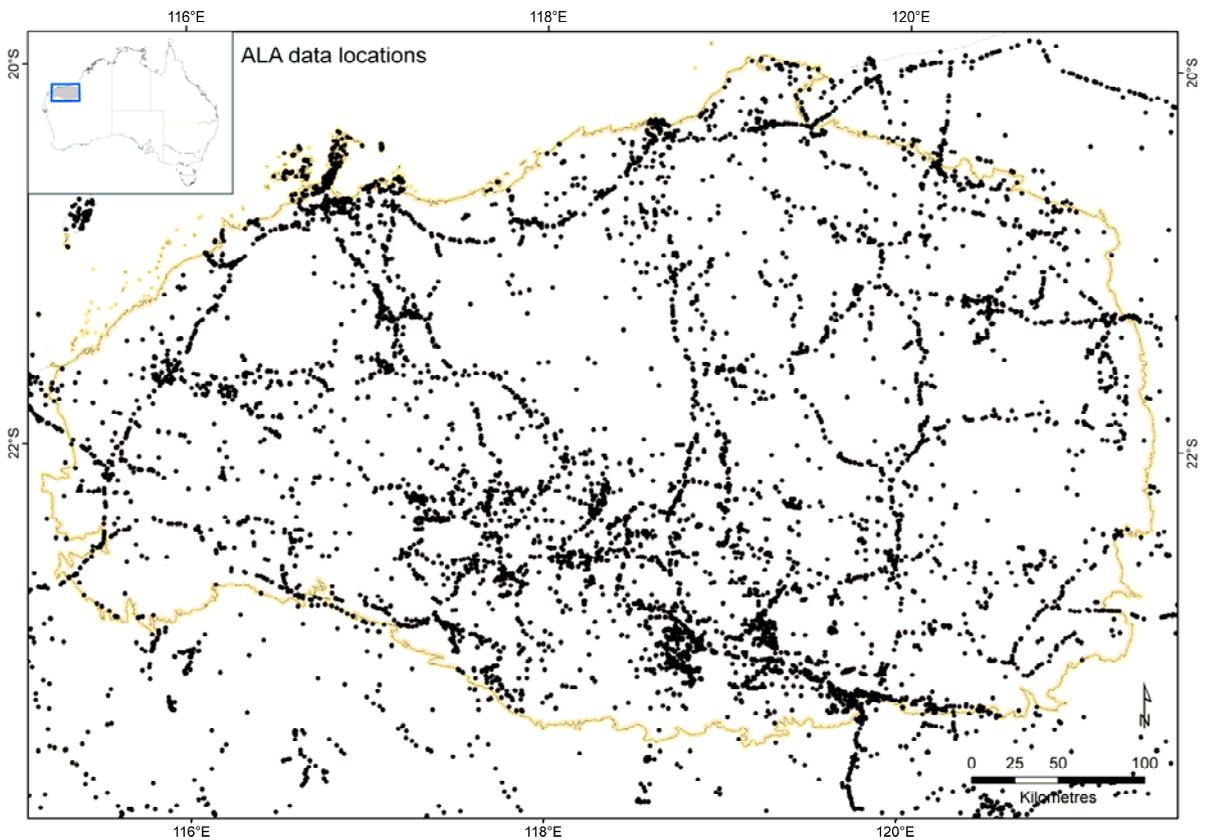
Figure 6. Locations of biological data sourced from the Atlas of Living Australia aggregation (www.ala.org.au).

Table 2. BHPBIO datasets assessed with biological records. A few minor changes were made to ensure uniform formats for collation.

DATASET NAME	DESCRIPTION
REG_FLORA_POPULATION	Estimate of tree species (2 species) broad numbers
FLORA_SAMPLESITE	Flora species
REG_SHORT_RANGE _ENDEMIC_FAUNA	Invertebrate point records
REG_SIGNIFICANT_FLORA	Point Vegetation data, with counts, etc
REG_STYGOFAUNA	Point data for Stygofauna all within bounds of AOI, with dates etc
REG_TROGLOFAUNA	Troglofauna within the area, abundance etc
REG_VERT_FAUNA_OBSERVATIONS	Point location of vertebrate fauna with full description
BARIMUNYA_CAMP DATA.	XBarCamp_Flora_ZN50_0911.csv XBarCamp_PrioFlo_ZN50_0911.csv
CALLAWA_WEST_ONSHORE_FLORA_SURVEY_FINAL.ZIP.TXT	Cal_FLORA_SAMPLE_SITE.csv, CAL_SIGN_FLORA_OB.csv
ONSHORE ENV_YARRIE.ZIP.TXT	On_shore_VERTEBRATE_FAUNA.csv
ONSHORE_MARILLANA_2012_02_12.ZIP.TEXT	xMAN_FLORA_SAMPLE_SITE.csv
MUDLARKVERTFAUNASURVEY_BHPDATABASE_15022013.XLSX	Vertebrate Fauna sample sites Significant Fauna Habitat
TANDANYA_COMBINED(2).XLS	Flora, Significant Flora
FRM-IEN-EMS-002	Flora, Significant Flora
AREACWESTVERTFAUNASURVEY_BHPDATABASE_15022013.XLS	vertebrate fauna
AREAC WEST TO TANDII	Flora, Significant Flora

2.3.2 DATA VETTING PROCEDURES

The process of data vetting involves three stages:

1. Match species names to a common list
2. Evaluate location accuracy
3. Evaluate geographic spread of biological groups

A summary of the requirements and a case study for assessing biological data suitability for community-level analyses is presented in Zerger *et al.* (2013). Our approach closely follows the procedures outlined in Chapter 6 of that report.

The Atlas of Living Australia also flags any issues of data quality that can be detected by various rules and algorithms. For example, the record of a terrestrial species occurring in the ocean will be flagged but the coordinates will not be altered. Taxonomic checks are performed against the National Species List. A suite of five reasonably independent terrestrial climate variables are used to detect environmental outliers among species records. Questionable data are flagged rather than omitted from the ALA aggregations. The data downloaded from the Atlas include these flags and so enable records with potential issues to be assessed and removed if they are likely to affect subsequent analyses.

For the generalised dissimilarity modelling to perform well, the biological group should contain a relatively large number of occurrence records, be spread across the four subregions and representative of the majority of environments in which species from that group are found. Ideally, the aggregate of records within a grid cell should approximate presence-absence sampling. The spatial grid for aggregation in this study is 9-second (approximating 250x250m). Large aggregations of data from multiple sources can begin to approach a presence-absence sample in some parts of the landscape.

2.3.3   SPATIAL ENVIRONMENTAL DATA

Spatial environmental data were compiled according to a general conceptual model of the relationship between biodiversity and its habitat (Williams et al. 2012). For terrestrial biodiversity this general model assumes responses to both physical and biological components of their environment (e.g., Mckenzie *et al.* 2000b; McKenzie *et al.* 2004; George *et al.* 2011). The physical environment can be described by facets of climate, regolith and landform and the biological environment may be described directly by vegetation patterns or indirectly by the physical environmental correlates of vegetation patterns (but see Leibold *et al.* 2004; Leibold *et al.* 2010 for reviews of other influences on species turnover). For the Pilbara region, we targeted collation of spatial environmental data within these broad classes or their proxies (Appendix B).

2.4 Results

2.4.1   MATCHED NAMES

A common list of taxa names was developed from the three data sources. The names in this list were compared with the Australian Plant Census (APC) and the Australian Faunal Directory (AFD) using tools available though the Atlas of Living Australia (http://www.ala.org.au/spatial-portal-help/getlsid-spreadsheet-macro/). Taxonomic rank was added and mismatched species names were corrected (synonyms and spelling errors) to enable the data sources to be merged. Unspecified species (c.f., x and sp.) were matched to genus or above levels. A list of the feral animals of Australia was used to identify and remove those species from the analysis. A digital list of introduced plants for Australia was extracted from Groves *et al.* (2003). This list comprises about the same number (2700) as the established species in the review by Randall (2007) and enabled listed plant species to be removed prior to analysis.

Most of the invertebrate records in the BHPBIO data sources were listed at the genus level (Table 3) and therefore could not be included in the analysis. The invertebrate records from the ALA aggregations were also limited in their suitability for analysis (Table 4). The terrestrial invertebrate survey data from DPaW includes a number of unnamed invertebrate species but these were consistently associated with survey sites. Therefore, only the DPaW invertebrate data (Durrant *et al.* 2009; Guthrie *et al.* 2009; Heterick *et al.* 2009; Volschenk *et al.* 2009) could be used in subsequent analyses for that biological group. Because the DPaW survey of scorpions (Volschenk *et al.* 2009) was incomplete in this data (N McKenzie pers. comm.), only the spiders (Order Araneae), beetles (Order Coleoptera) and ants (Order Hymenoptera) were included.

Data were compiled from across all three data sources for the three vertebrate groups – mammals, birds and reptiles – and for vascular plants. While the majority of the vascular plant survey data derived from BHPBIO, the ALA contributed data more broadly across the region of interest (Appendix A). From DPaW, the riparian flora (M. Lyons pers. comm.) contributed records of vascular plants (Table 4). Other sources of survey data from DPaW include mammals (Gibson & McKenzie 2009), reptiles (Doughty *et al.* 2009) and birds (Burbidge *et al.* 2009).

Data sourced from the ALA aggregation (as at April 2013) included records from the Western Australian Museum and the Western Australian Herbarium as well as other contributing museums and herbaria from around Australia (see http://www.ala.org.au/about-the-atlas/atlas-background/atlas-partners/) and overseas (aggregated by the ALA via partners of the Global Biodiversity Information Facility, http://www.gbif.org/).

Table 3. Number of species and genera associated with the aggregated BHPBIO data. Numbers of records and taxa are for unique combinations. Locations at this stage are unvetted and may extend beyond the study region. Taxa may include introduced species.

BIOLOGICAL GROUP	RECORDS BY SPECIES	SPECIES	RECORDS BY GENUS	GENERA
Class Mammalia (mammals)	3352	62	3373	43
Class Aves (birds)	4782	212	4496	137
Class Reptilia (reptiles)	2560	130	2111	52
Order Araneae (spiders)	1	1	367	19
Order Carabidae (beetles)	0	0	1	1
Order Gastropoda (slugs, snails)	48	5	95	6
Order Hemiptera (bugs)	0	0	1	1
Order Hymenoptera (ants, bees, wasps)	0	0	15	2
Order Scopiones (true scorpions)	8	2	99	2
Class Equisetopsida (Vascular plants[1])	72,851	1175	61,594	321

Table 4. The vetted number of records of named species for the biological groups selected for analysis. Numbers of records (species by location) and species are for unique combinations (excludes introduced species).

DATA SOURCE	BIOLOGICAL GROUP	RECORDS BY SPECIES	SPECIES
BHP	Class Mammalia (mammals)	3184	59
	Class Aves (birds)	4813	220
	Class Reptilia (reptiles)	2560	142
	Order Araneae (spiders)	1	1
	Order Carabidae (beetles)	0	0
	Order Hymenoptera (ants, bees, wasps)	0	0
	Class Equisetopsida (vascular plants[1])	64,237	1289
DPaW	Class Mammalia (mammals)	963	19
	Class Aves (birds)	5558	127
	Class Reptilia (reptiles)	3100	107
	Order Araneae (spiders)	3662	343
	Order Carabidae (beetles)	3442	421
	Order Hymenoptera (ants, bees, wasps)	5016	241
	Class Equisetopsida (vascular plants[1])	2869	477
ALA	Class Mammalia (mammals)	1681	41
	Class Aves (birds)	58,334	260
	Class Reptilia (reptiles)	112	49
	Order Araneae (spiders)	775	132
	Order Carabidae (beetles)	8	8
	Order Hymenoptera (ants, bees, wasps)	24	14
	Class Equisetopsida (vascular plants[1])	21,665	1723

1. Vascular plants comprise clubmosses, horsetails, ferns, gymnosperms and angiosperms

2.4.2 LOCATIONAL ACCURACY

We used data from the ALA that were not flagged with data quality issues. All data sets were converted to latitude and longitude in decimal degrees in the world geodetic datum, WGS 1984, if not already in that spatial reference system. Bulk conversion of UTM gridded data was completed using an R script via the Redfern equation (http://www.ga.gov.au/nmd/geodesy/datums/redfearn_grid_to_geo.jsp).

Records were excluded from the analysis on the basis of:

2. Missing or inaccurate co-ordinate systems data (accuracy if specified, < 1000m)
3. Defined outside the area of interest (> -40 Latitude, >140 longitude)
4. Not uniquely identified at the species-level of classification

The resulting species by locations data were then aggregated to a 9-second grid of latitude and longitude, representing the spatial resolution of the analysis (Table 5).

Table 5. Number of aggregated analysis sites (9 sec grid of longitude and latitude).

BIOLOGICAL GROUP	ALL RECORDS	ALL LOCATIONS	AGGREGATED RECORDS	AGGREGATED SITES
Class Mammalia (mammals)	5825	3781	3794	2318
Class Aves (birds)	68,701	6384	55,158	5271
Class Reptilia (reptiles)	5772	1120	5364	915
Invertebrates (ants, beetles, spiders)	12,239	297	12,092	296
Class Equisetopsida (vascular plants[1])	88,768	16,493	77,447	10,475

1. Vascular plants comprise clubmosses, horsetails, ferns, gymnosperms and angiosperms

2.4.3 GEOGRAPHIC SPREAD

Geographic spread of each short listed group was examined by plotting the distribution of points for the four cases: all records, BHPBIO, DPaW, ALA. Maps for each group are presented in Appendix A . Among these, only the groups listed in Table 5 were considered suitable for further analysis.

To improve the comprehensiveness of sampling when using aggregated data, the minimum species richness at a site provides a surrogate for improving data quality. Locations with only one or two occurrences of a species in a biological group are typical of *ad hoc* observations and may also be indicative of spatial location errors. Information about the minimum number of species recorded at a site following comprehensive survey (e.g., DPaW data) also provides a useful measure of the expected number of species at a site. Generalised dissimilarity modelling (GDM) is robust to sparsely sampled comprehensive and representative survey data. The GDM method also performs reasonably well with data aggregated from different sources that have been assessed fit for purpose and filtered to improve data quality. We therefore tested different species by site richness thresholds in preliminary generalised dissimilarity models and selected thresholds applicable to the respective biological group. There was no need to filter the terrestrial invertebrate data which derives from a comprehensive survey, other than to remove scorpions from the analysis because this group was under sampled.

The resulting sets of locations for each of the biological groups and their relative richness are shown in Figure 7 to Figure 11, with richness filters applied as shown in Table 6. The more intensive sampling associated with mining tenement areas is indicative of observer bias. In Section 3.1 we evaluate how representative this sampling is in the context of the data used in the GDM model for each biological group.

Table 6. Data compiled for analysis of compositional turnover (aggregated 9 sec grids of longitude and latitude).

BIOLOGICAL GROUP	NUMBER OF RECORDS	NUMBER OF LOCATIONS	NUMBER OF SPECIES	RICHNESS FILTER APPLIED
Class Mammalia (mammals)	4023	2443	72	> 2
Class Aves (birds)	48,205	5364	304	> 10
Class Reptilia (reptiles)	5466	931	185	> 2
Invertebrates (ants, beetles, spiders)	12,092	296	1005	None
Class Equisetopsida (vascular plants[1])	90,019	13,462	2515	> 10
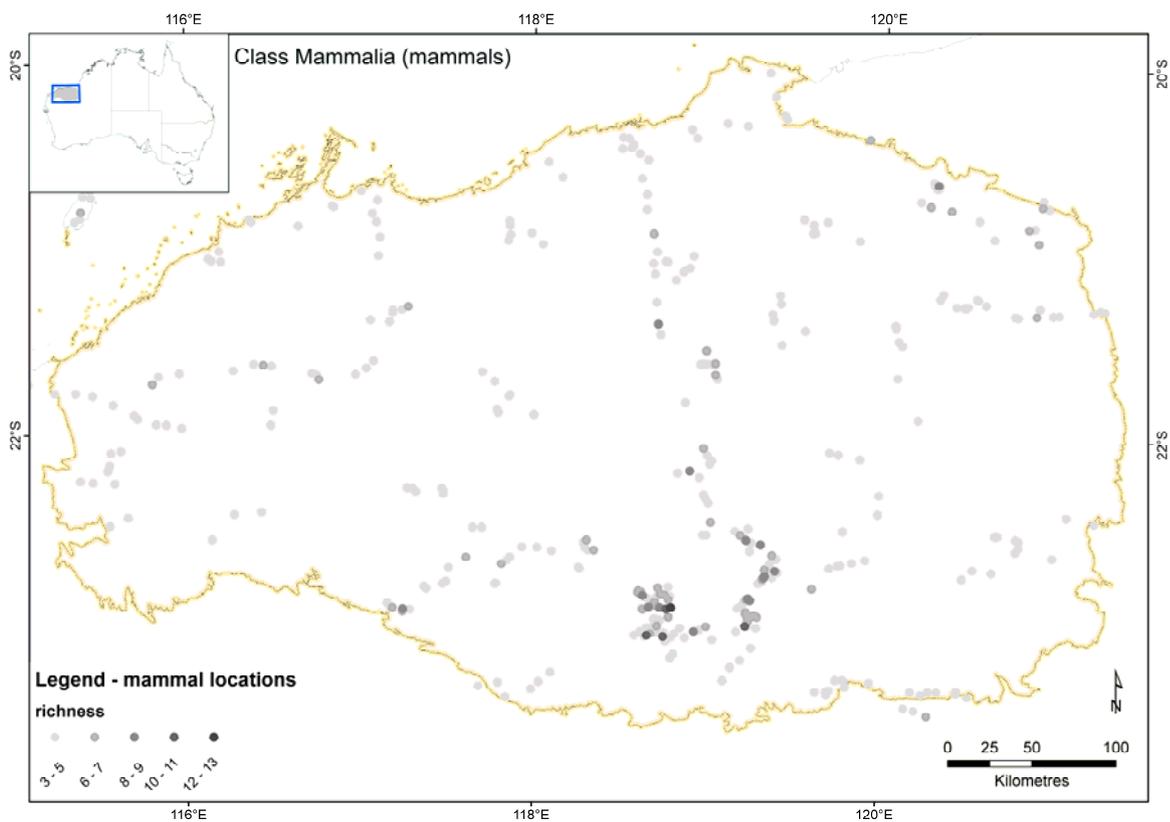
Figure 7. Mammal locations used in the GDM analysis showing species richness within 9-second grids for grid-sites with more than two species.
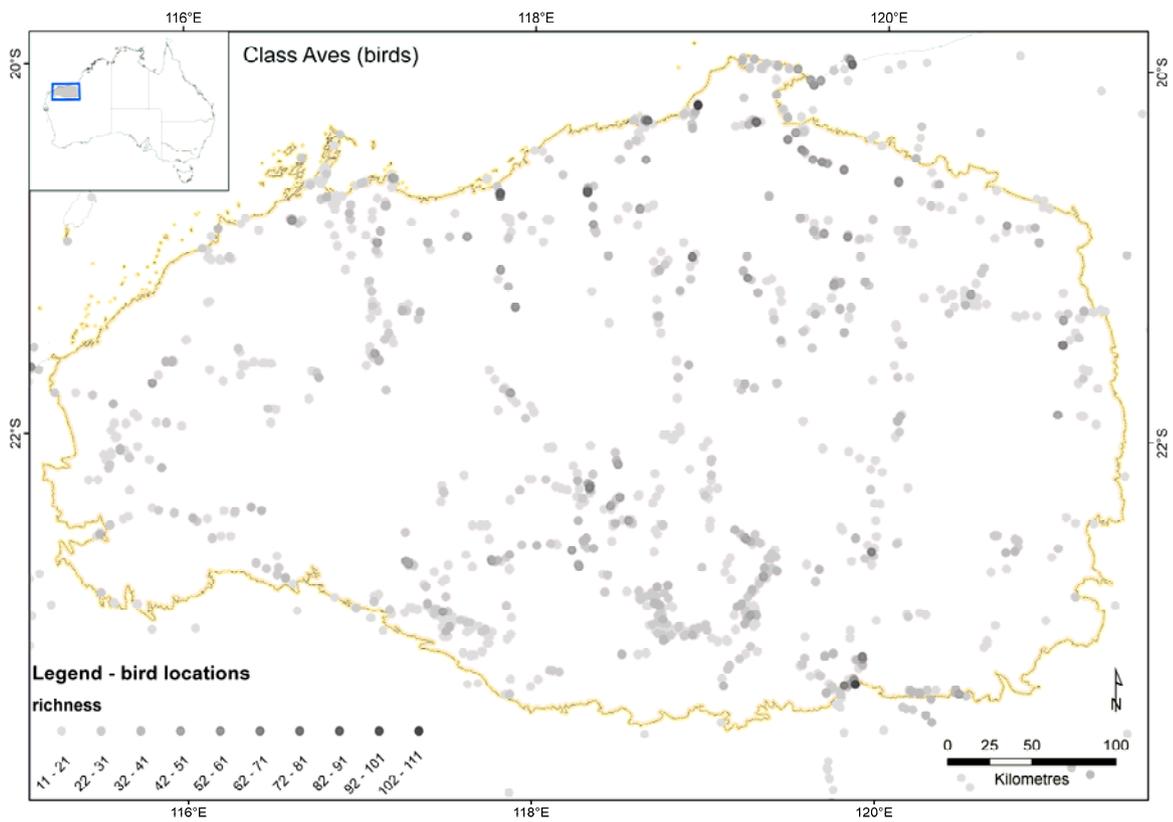
Figure 8. Bird locations used in the GDM analysis showing species richness within 9-second grids for grid-sites with more than 10 species.
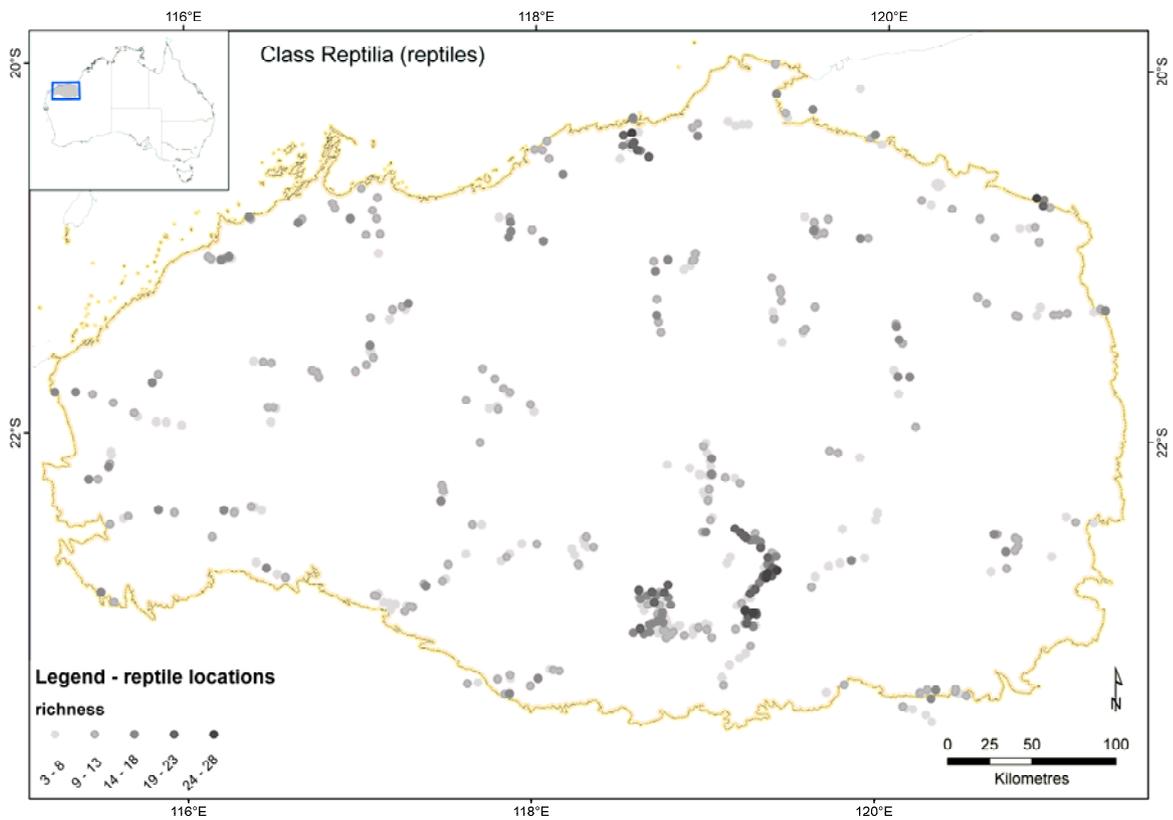
Figure 9. Reptile locations used in the GDM analysis showing species richness within 9-second grids for grid-sites with more than two species.
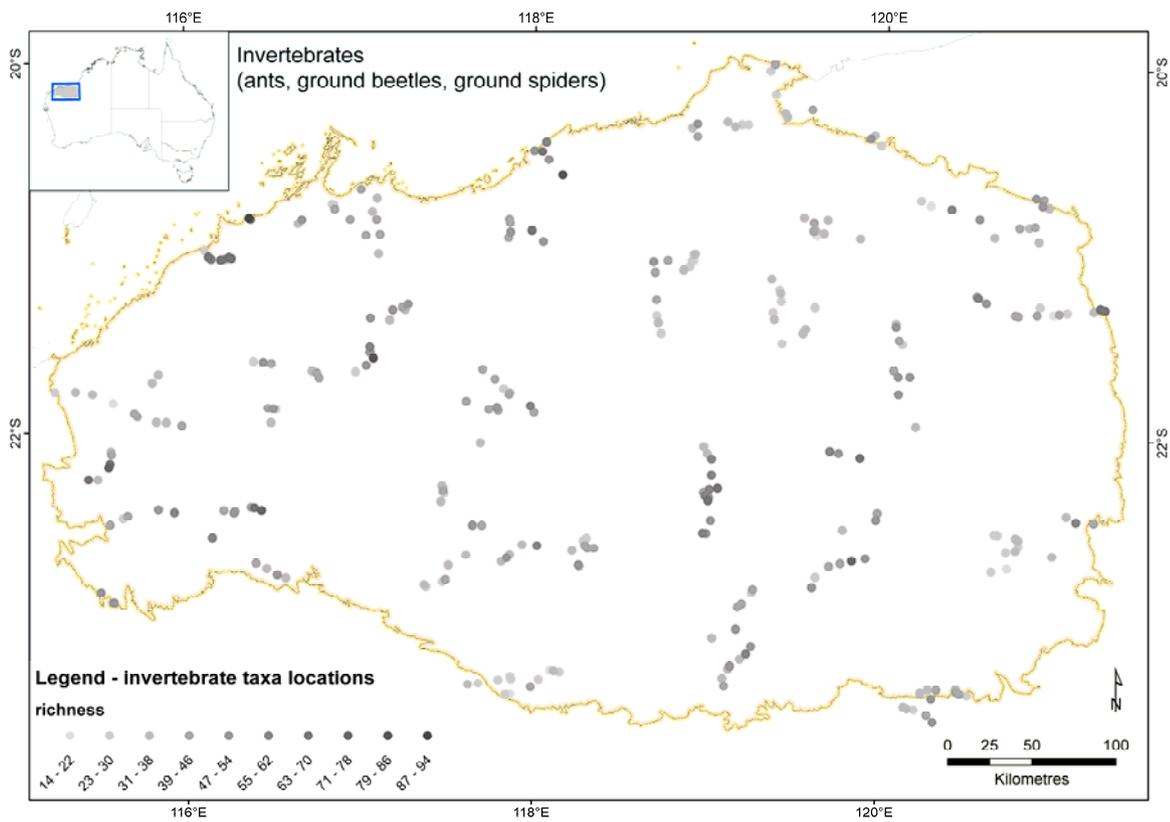
Figure 10. Terrestrial invertebrate locations used in the GDM analysis showing species richness within 9-second grids. No richness filtering was applied as the sample derives from comprehensive surveys.
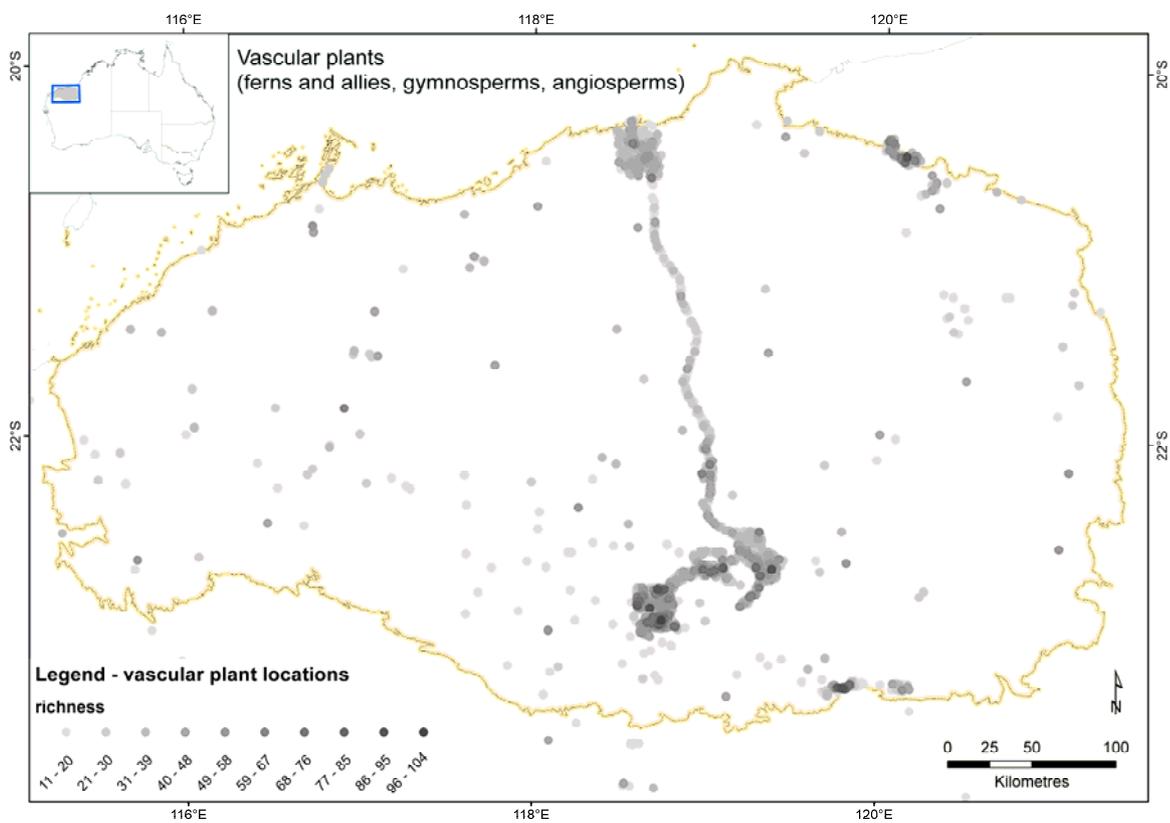
Figure 11. Vascular plant locations used in the GDM analysis showing species richness within 9-second grids for grid-sites with more than 10 species.

2.4.4 COMPOSITIONAL DATA FOR SPECIES RICHNESS MODELLING

While the biological data requirements for generalised dissimilarity modelling can include *ad hoc* surveys, with some filtering of richness to approximate comprehensive surveys, the modelling of species richness requires greater rigour and only data from comprehensive surveys can be included. Suitable comprehensive survey data were sourced from DPaW (invertebrates and vertebrates) and BHPBIO (plants and vertebrates). From the BHPBIO data, only those sources with a specified sampling method (e.g., trap or bird survey), or those with an associated quadrat number or code indicating that a systematic rather than an *ad hoc* survey was conducted, were included (sources listed in Table 7). Details about the selection of DPaW sites for survey are provided in the published survey reports (McKenzie *et al.* 2009; George *et al.* 2011). The vascular plant surveys in particular are concentrated around the BHPBIO tenements (Figure 12). While we included the DPaW surveys of riparian vascular plants in this analysis, the more terrestrial survey data were unavailable because species identifications are incomplete (Stephen Van Leeuwen personal communication). An estimate of statistical uncertainty derived from the models of species richness in Section 3.2 provides an indication of spatial sampling bias and potential areas where additional sampling would improve the robustness of these data for modelling.

Table 7. BHPBIO comprehensive survey data assessed suitable for the analysis of species richness.

GROUP	FILE NAME / SURVEY
Plants	FLORA_SAMPLESITE
	BARIMUNYA_CAMP DATA.
	CALLAWA_WEST_ONSHORE_FLORA_SURVEY_FINAL.ZIP.TXT
	ONSHORE_MARILLANA_2012_02_12.ZIP.TEXT
	TANDANYA_COMBINED(2).XLS
	TERRESTRIAL FLORA (RIPARIAN)
Vertebrates	REG_VERT_FAUNA_OBSERVATIONS
	MUDLARKVERTFAUNASURVEY_BHPDATABASE_15022013.XLSX
	AREACWESTVERTFAUNASURVEY_BHPDATABASE_15022013.XLS
	TERRESTRIAL FAUNA

Table 8. Comprehensive survey data compiled for analysis of species richness (aggregated 9 sec grids of longitude and latitude).

BIOLOGICAL GROUP	RECORDS	LOCATIONS	SPECIES
Class Mammalia (mammals)	2455	1270	69
Class Aves (birds)	7756	551	283
Class Reptilia (reptiles)	4802	565	209
Invertebrates (terrestrial only)	12,092	296	1005
Class Equisetopsida (vascular plants[1])	58,171	2501	1431
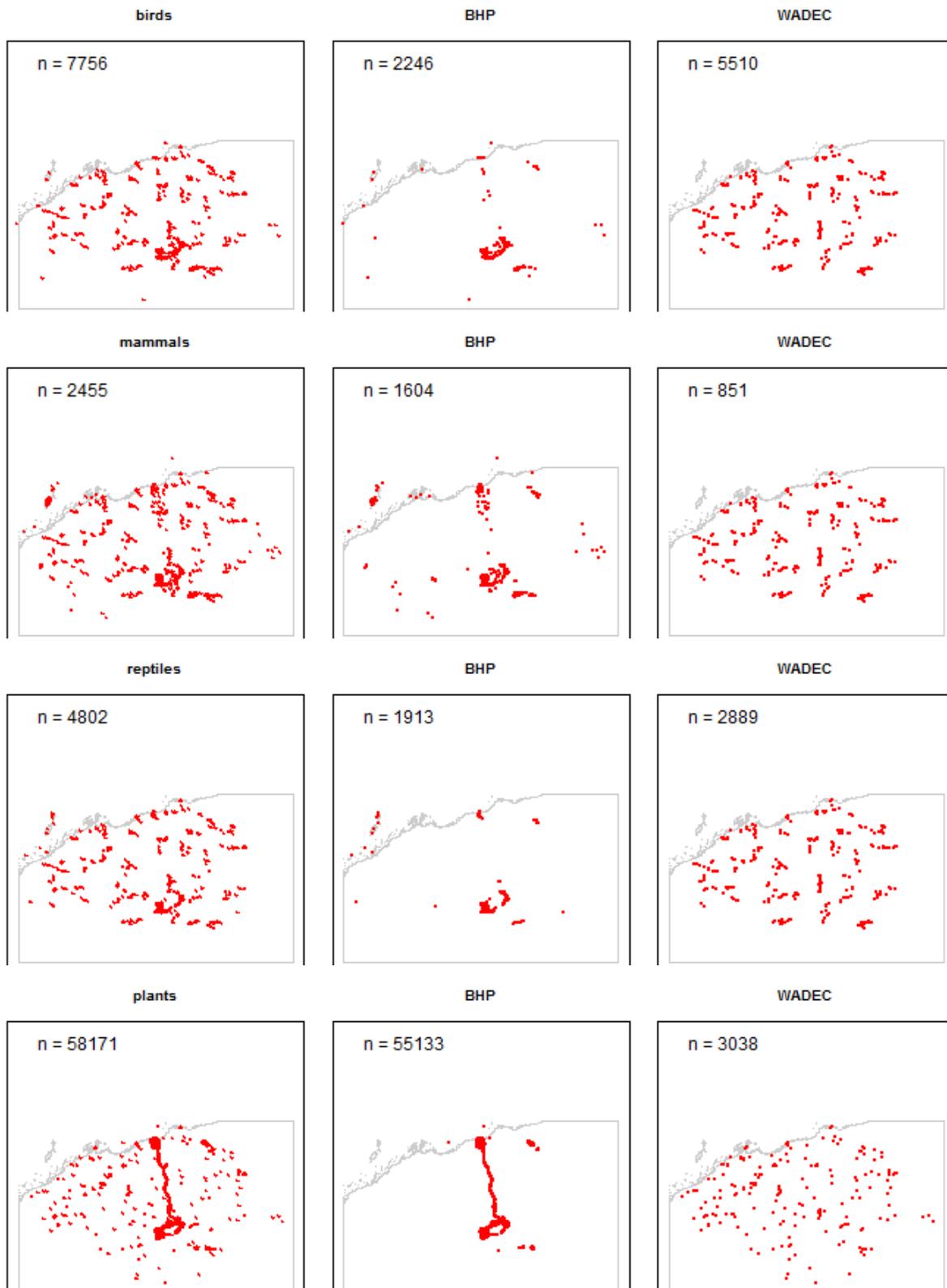
Figure 12. Geographic spread of comprehensive survey data for (from top) birds, mammals, reptiles and plants. Invertebrate data locations are shown in Figure 10.

2.5 Discussion

We conducted an assessment of all available sources of biological data and their integration for use in community-level modelling. This integration requires that all species names be related by a common nomenclature and at a common taxonomic rank (e.g., species rank). A key objective was to assess the fitness-for-use of the biological data compiled by BHPBIO and to recommend corrective actions. Through this section, we presented the framework and methodology used to assess the available data. Tools that are publicly available through the ALA enabled common nomenclatures to be established for named flora and fauna. The data compiled by BHPBIO are well managed and integrated, covering a complex array of survey methods and purposes with sufficient attribute detail to provide guidance about the source and quality of the data that could be investigated further by referring to individual consultant reports, as needed. This effective data management made the process of aggregating and filtering data across the three primary sources a relatively straight forward process. However, a systematic assessment and annotation with feedback record by record, dataset by dataset was beyond the scope of this work.

The data that were found to be of least use were the invertebrate group which were commonly identified only to genus level. We therefore relied upon the invertebrate data from Pilbara biological surveys for this analysis (George *et al.* 2011). A genus level dataset was subsequently generated to evaluate whether this taxonomic rank, with substantial BHPBIO data included, could usefully inform a compositional turnover model.

The data requirements for compositional turnover and richness modelling are different. Although both methods benefit the most from comprehensive and representative survey data generated using a consistent site-based survey methodology, such data are typically only sparsely available and must be supplemented from other sources and data aggregations. Generalised dissimilarity modelling of compositional turnover is relatively robust to variation in data quality (sampling comprehensiveness) if this is complemented by a geographic (and therefore environmental) spread of data over the region of interest. However, the modelling of species richness requires greater rigour in the selection of survey data suitable for this analysis. In this respect, we used only those records from the BHPBIO data, for which a specified sampling method was provided (e.g., trap or bird survey, quadrat number, etc) to infer a systematic rather than *ad hoc* or targeted survey. This substantially reduced the sample of data available for this analysis, and likely included a mixture of survey type and therefore sampling comprehensiveness that may limit the capacity to interpolate richness across the region.

3 Biodiversity model fitting

3.1 Generalised dissimilarity modelling of species composition

3.1.1 INTRODUCTION

Generalized dissimilarity modelling (GDM) is a statistical technique for modelling the compositional dissimilarity (Czekanowski 1932) between pairs of geographical locations, for a given biological group (e.g., using species records), as a non-linear multivariate function of environmental differences between these locations (Ferrier *et al.* 2007). The compositional dissimilarity between a given pair of locations can be thought of as the proportion of species occurring at one location that do not occur at the other location (averaged across the two locations) – ranging from '0' if the two locations have exactly the same species through to '1' if they have no species in common. GDM effectively weights and transforms the environmental variables such that distances between locations in this transformed multidimensional environmental space now correlate, as closely as possible, with the observed biological compositional dissimilarities between these same locations (see Ferrier *et al*. 2007 for full explanation).

Once a GDM model has been fitted to the biological data from the sampled locations, it can be used to predict compositional dissimilarity values for sites lacking biological data, based purely on their mapped environmental attributes. Although factors other than local environmental controls can influence species turnover, including metacommunity processes such as competitive niche-differentiation, historical biogeographic events and disturbance heterogeneity (e.g., Leibold *et al.* 2004; Urban 2004; Armstrong 2005; Leibold *et al.* 2010), this predictive capacity provides, in turn, a foundation for performing various subsequent spatial analyses – such as biodiversity significance assessments and survey gap analyses.

In this section we describe the application of GDM to develop models of species compositional turnover for the Pilbara, using the data outlined in Table 6. The results are then used to assess gaps in the biological data used to fit the models as an indicator of spatial uncertainty in the model outputs and therefore uncertainty in subsequent biodiversity assessments.

The analysis domain includes the Pilbara bioregion and surrounding environmental gradients east to west and north to south (Figure 2). The spatial unit is a 9-second geographic grid (approximately 250m) in GDA94 (synonymous with WGS84 at this resolution), approximating the size of the site for field survey sampling; ca. 1ha for most zoological groups (McKenzie *et al.* 2009).

3.1.2 METHODS

We used a recently revised version of the .NET software (Manion 2013) to fit generalised dissimilarity models (GDM) to the aggregated biodiversity data (Figure 13). The .NET software requires three inputs: 1) species location records (biological data), with provision for different input format options, 2) spatial analysis domain which defines the grid extent and the location of data/no-data cells; and 3) environmental predictor data as spatial layers in ESRI binary float-grid format (*.flt). The lists of species recorded at each of two sites for all site-pairs are used to calculate compositional dissimilarity, the response variable in GDM (see equation 3 in Rosauer *et al.* 2013). The response variable can be weighted by estimates of sampling intensity (or detectability) associated with each site. We used the sum of the number of species recorded at each of a pair of sites, irrespective of any species in common, as the input to the weight function. This weight variable acts as a proxy for potential under-sampling of species between sites. The outputs of the model include a set of scaled environmental layers that have been selected and transformed by the coefficients of the fitted model.

The variable selection strategy followed the stage-wise process outlined by Williams *et al.* (2012) where different groups of correlated variables are initially tested for redundancy before combining and removing relatively insignificant variables using a backward elimination procedure that tests the contribution of each variable. Preliminary models explored the effectiveness of a large number of candidate predictors grouped into climate (27), regolith (35), landform (15), hydrological and land cover (6) (detailed in Appendix B). The more marginal of the remaining active predictors (usually around 40) were successively removed using a stopping criterion of 0.05% partial deviance explained. This value was determined to be a reasonable tradeoff in parsimony between the number of predictors included in the model and cumulative reduction in deviance explained by the model.
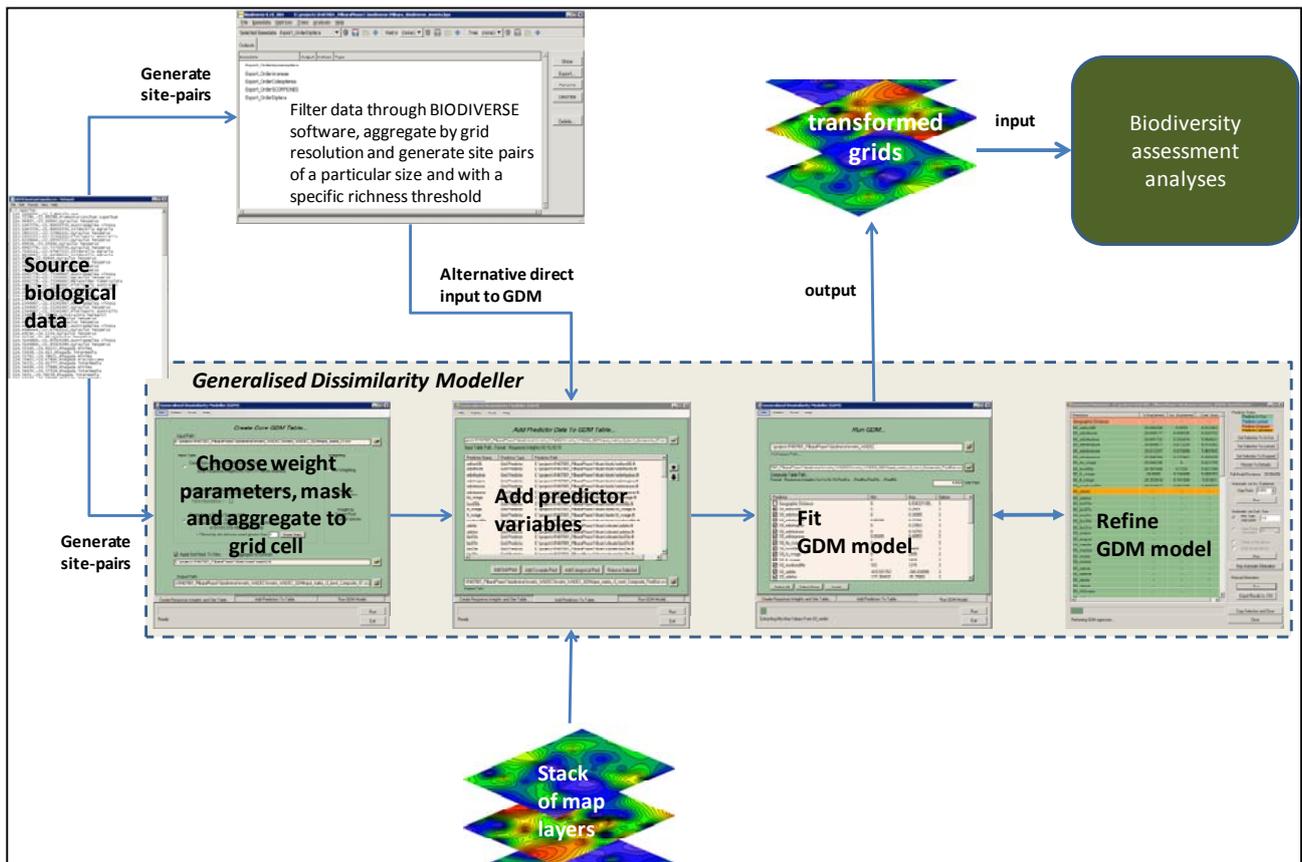
Figure 13. Schematic showing steps in the development of GDM outputs. For large datasets, *Biodiverse* software (Laffan *et al.* 2010) is used with a custom tool (Rosauer 2009) for sampling site pairs from millions or billions of possible combinations.

To visualise the spatial structure of compositional turnover, we used a sample of 10,000 grid-cells evenly spread in geographic space across the study area to derive a 300 group agglomerative hierarchical classification (the 'UPGMA' algorithm) of predicted Sorenson dissimilarity based on the GDM-scaled environmental space (Ferrier *et al.* 2007). Class colours represent class compositional similarity derived from their Red-Green-Blue assignment to the first three axes of a multi-dimensional scaling of predicted similarity (Belbin *et al.* 1983).

No formal technique has yet been developed for directly estimating, and mapping, the level of uncertainty associated with predictions from GDM models. We therefore employed sampling density in environmental space as a relative indicator of the level of certainty associated with predictions of compositional turnover derived from the GDMs fitted within this environmental space. The density of biological sampling relative to a given grid-cell (i) is calculated as a weighted proportion of grid-cells sampled within the region (for the

biological group of interest), with weights set equal to the expected compositional similarity (= 1-dissimilarity) between cell i and each other cell j based on the fitted GDM for that biological group:

$$BD_i = \frac{\sum_{j=1}^{n} s_{ij} b_j}{\sum_{j=1}^{n} s_{ij}}$$

where b_j = 1 if grid-cell j has been surveyed for the biological group of interest, or 0 if it has not been surveyed for that group. This is necessarily a small proportion with a long tailed distribution. The relative uncertainty associated with the GDM model has therefore been classified in the range 0 to 1% coverage, although a few cells may have larger values.

3.1.3 RESULTS

Variation in the frequency distribution of observed dissimilarity data for each biological group reflects differences in sampling across the study region (Figure 14). Histograms that are strongly skewed toward dissimilarity=1, as is the case for reptiles and mammals, are indicative of data dominated by presence-only samples due to under-sampling. The more symmetric distribution of the response histograms for invertebrates and birds is more like a presence/absence sample. Sites with less than 10 recorded species were removed from the bird and vascular plant data prior to the calculation of site-pair dissimilarity, to improve data quality for use in GDM. The response variable for terrestrial invertebrates derives from systematic survey data (McKenzie *et al.* 2009) and the shape of the histogram is indicative of the distribution of a true presence-absence sample.

Summary statistics for the five fitted models are given in Table 9 and the model fit is shown in Appendix C (see Figure 43, 86). Of the more than 80 candidate environmental predictors, 65 were included in at least one of the five models (Appendix C , Figure 44 and Figure 45, page 87-88).

Table 9: Summary statistics for the five fitted and refined GDM models.

BIOLOGICAL GROUP	SITE-PAIRS	INTERCEPT[2]	% DEVIANCE EXPLAINED[3]	CONTRIBUTION OF ALL PREDICTORS TO TURNOVER[4]	NUMBER OF PREDICTORS	MAIN PREDICTOR[5]
Mammals[1]	81,403	0.42	17.64	10.17	20	RH2MAX
Birds[1]	121,662[*]	0.56	25.70	5.42	25	BIO30
Reptiles[1]	114,003	0.76	16.92	6.73	25	FORST06
Invertebrates	43,660	0.86	28.93	5.64	32	PC1
Vascular plants[1]	1,654,007[*]	1.03	35.37	14.56	23	RADNI

* randomly sub-sampled;
1. Modelled response weighted by the total number of species summed across the two sites in a pair.
2. The Intercept is a term in the linear regression model. A lower intercept, trending toward zero, implies a better fitting model.
3. % deviance explained refers to the amount of variation in the response data that is explained by the explanatory variables and can be approximated to an r^2 in a linear regression. A higher deviance explained implies a better fitting model, but a lower deviance explained may still be a credible model applied to "noisy" data.
4. The contribution of all predictors to turnover is the sum of the predictor coefficients in the model. Each predictor is defined by at least three terms (I-splines) to account for non-linearity in the relationship with the response. The sum of the coefficients provides an overall indicator of the amount of turnover predicted by the model – larger values imply higher rates of compositional turnover.
5. RH2MAX – maximum month of relative humidity; BIO30 – lowest period of moisture index; FORST06 - extent of woody vegetation in 2006 within grid cell; PC1 - Principal component 1 of spectra of surficial soils; RADNI - minimum month of solar radiation. Further details are given in Appendix B .

In order to visualize the prediction of compositional dissimilarity (which applies to each pair of cells in the spatial grid) we used an unsupervised classification to assign each cell to a specified number of groups (e.g., 300 in this case). The groups are then assigned colours according to their predicted compositionally similarity using the full colour spectrum (see top images in Figure 15 to Figure 19). These images show areas that are relatively similar or different. For example, if two localities are depicted with a similar colour,

then they are predicted to have a similar composition biologically (relative to the specified number of groups), but if they have a very different colour (e.g., at opposite ends of the red-green-blue spectrum), then they are predicted to have a composition that is quite different to each other. The colours carry no meaning in terms of high or low, nor comparatively between biological groups. Areas of model extrapolation may appear with quite different predicted compositions (e.g., such as the 'pink' outlier cells in the image for mammals, Figure 15). This can result in more subdued colouring overall. Model extrapolation occurs where unique types of environment are not represented in the sample data. The sampling coverage across the range of environments in the study area helps identify places where model extrapolation is occurring and predictions are therefore less reliable (see lower images in Figure 15 to Figure 19).
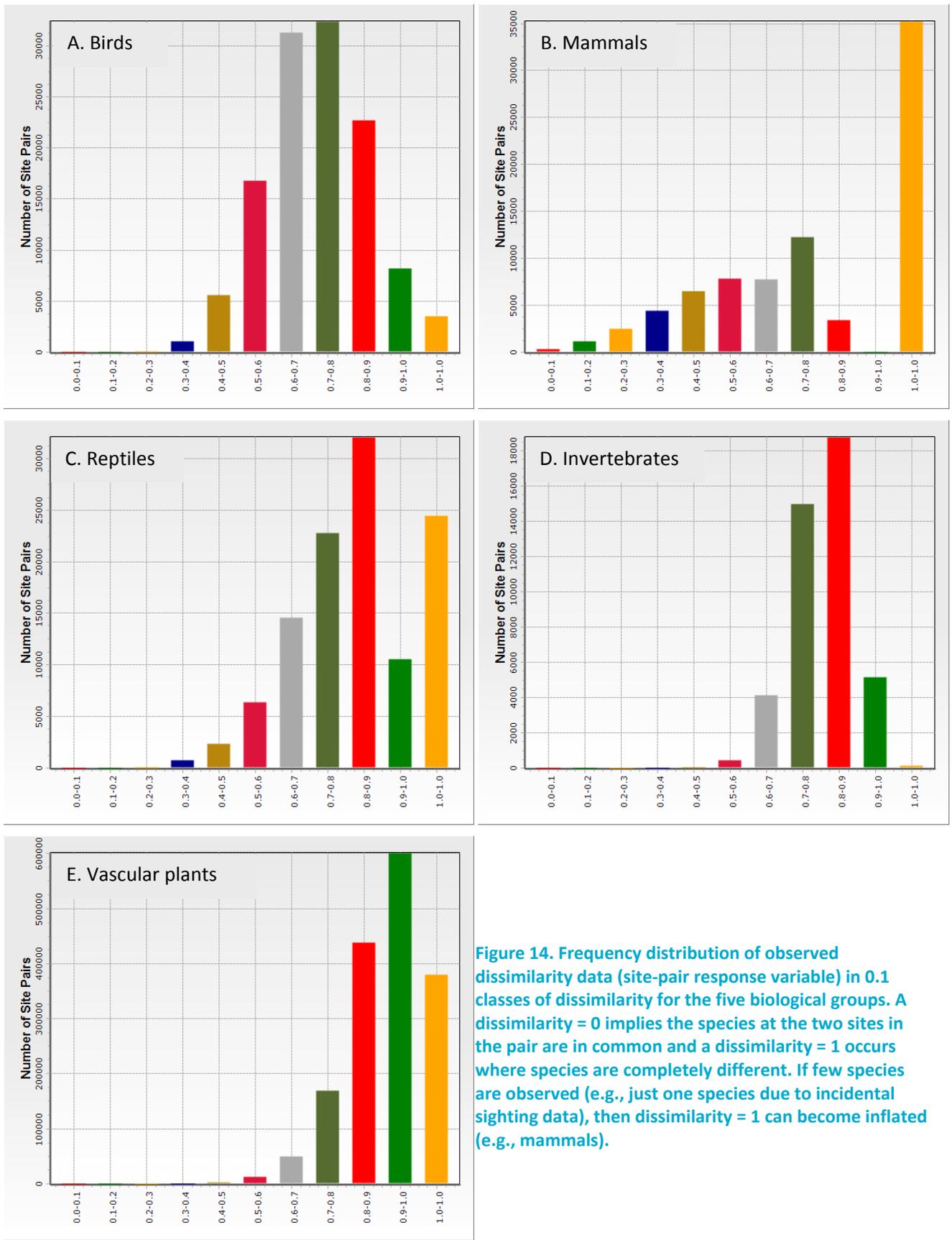
Figure 14. Frequency distribution of observed dissimilarity data (site-pair response variable) in 0.1 classes of dissimilarity for the five biological groups. A dissimilarity = 0 implies the species at the two sites in the pair are in common and a dissimilarity = 1 occurs where species are completely different. If few species are observed (e.g., just one species due to incidental sighting data), then dissimilarity = 1 can become inflated (e.g., mammals).
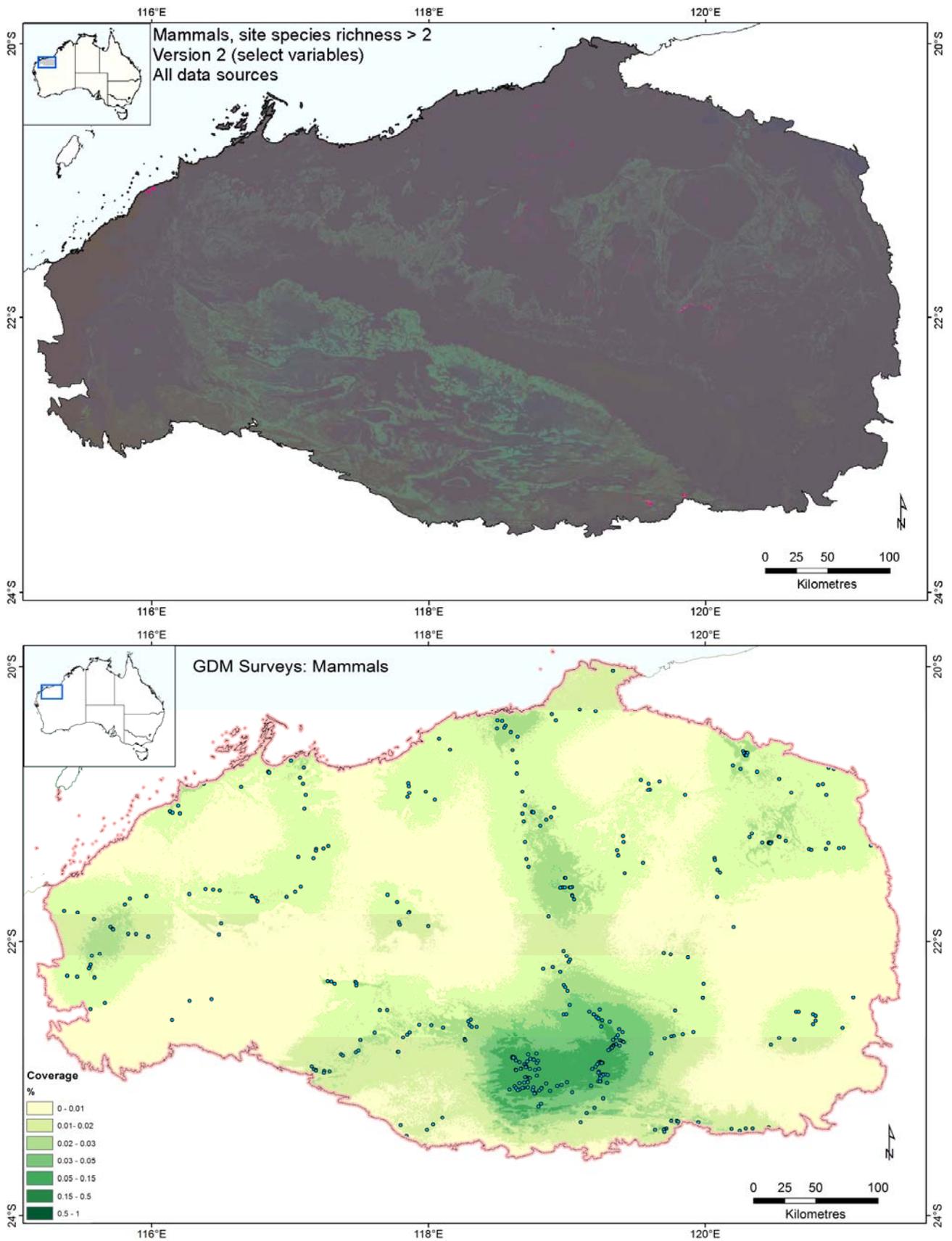
Figure 15. Classification of compositional turnover (upper) and estimated prediction uncertainty based on sampling coverage within GDM-scaled environmental space (lower) for mammals, shown for the Pilbara bioregion only.
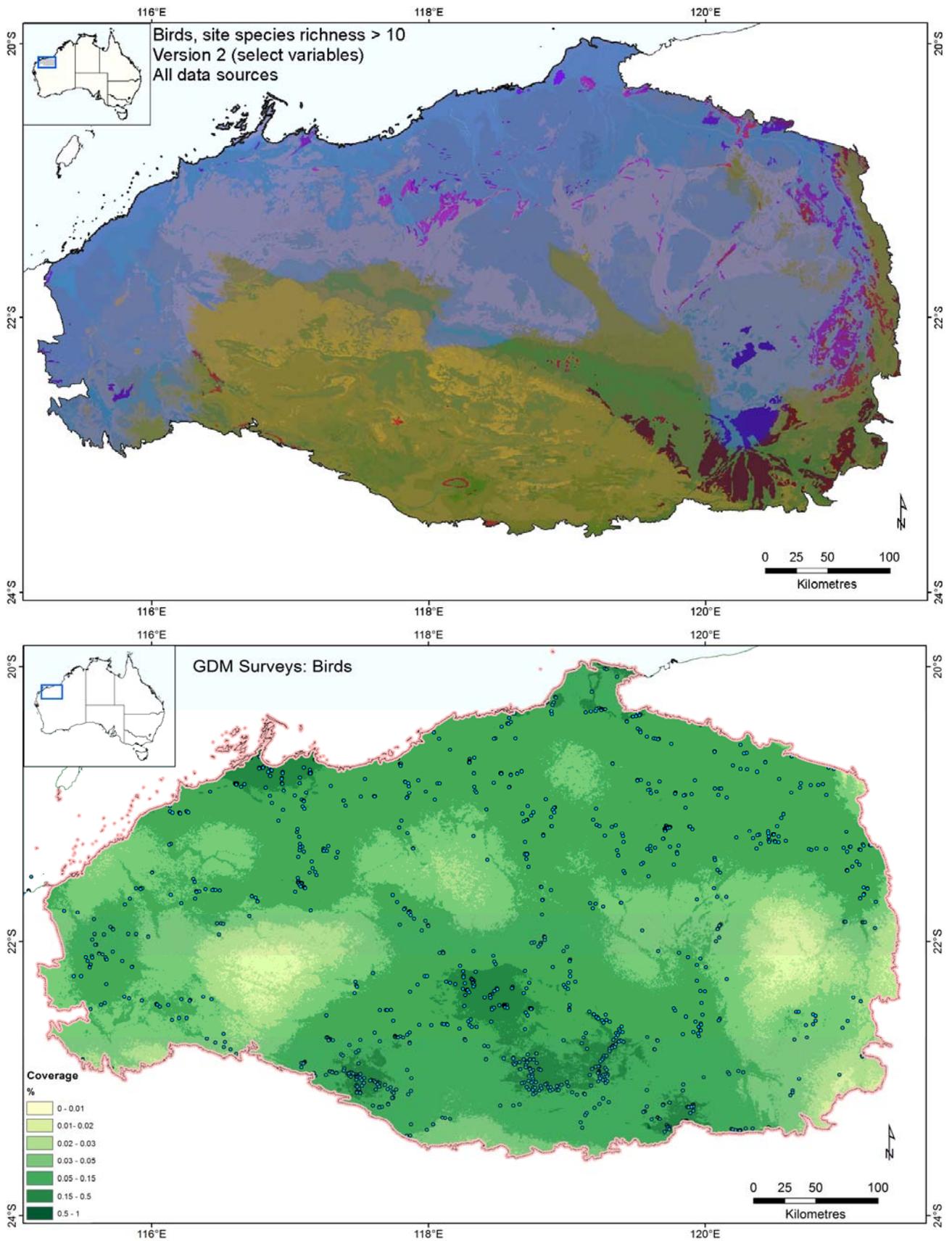
Figure 16. Classification of compositional turnover (upper) and estimated prediction uncertainty based on sampling coverage within GDM-scaled environmental space (lower) for birds, shown for the Pilbara bioregion only.
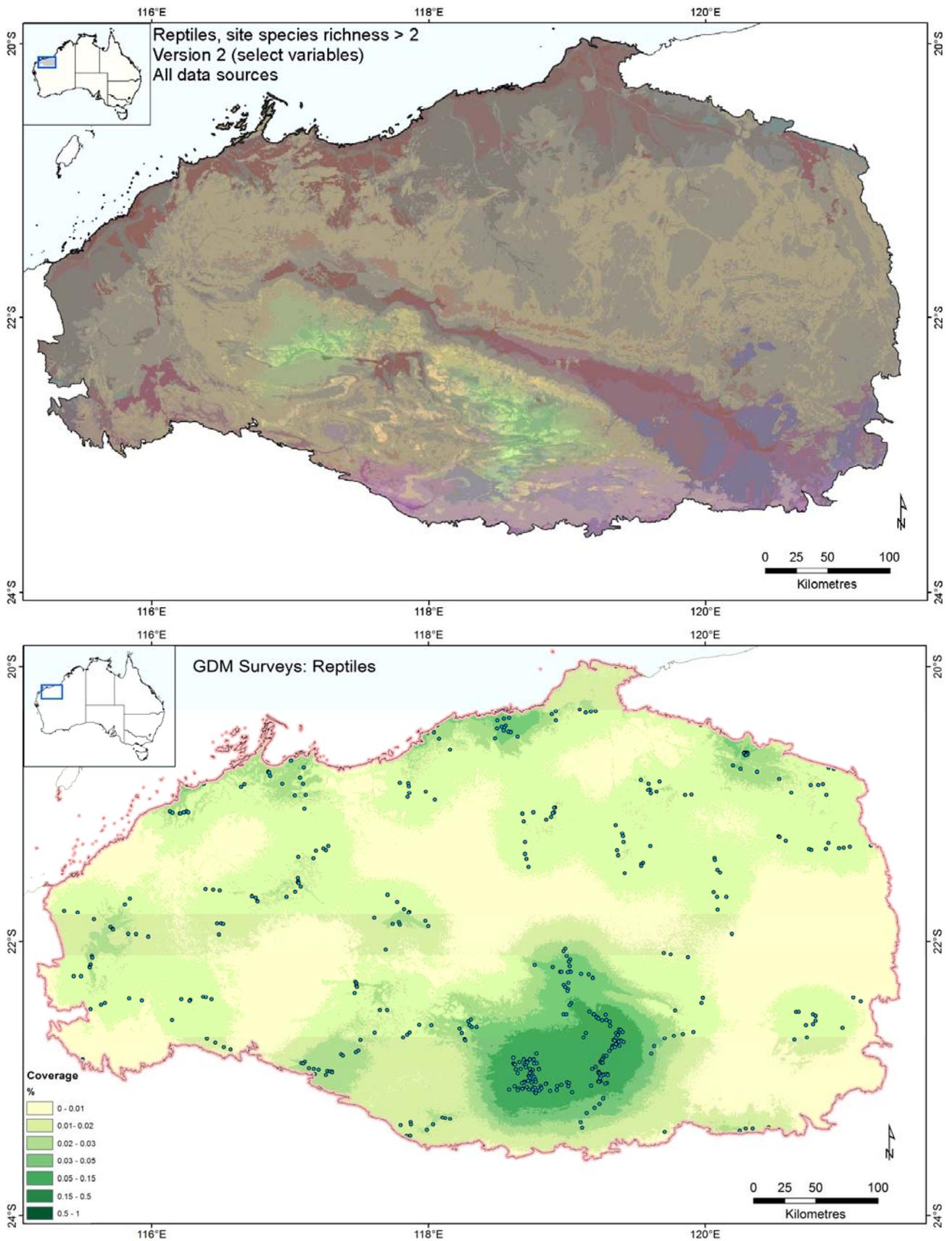
Figure 17. Classification of compositional turnover (upper) and estimated prediction uncertainty based on sampling coverage within GDM-scaled environmental space (lower) for reptiles, shown for the Pilbara bioregion only.
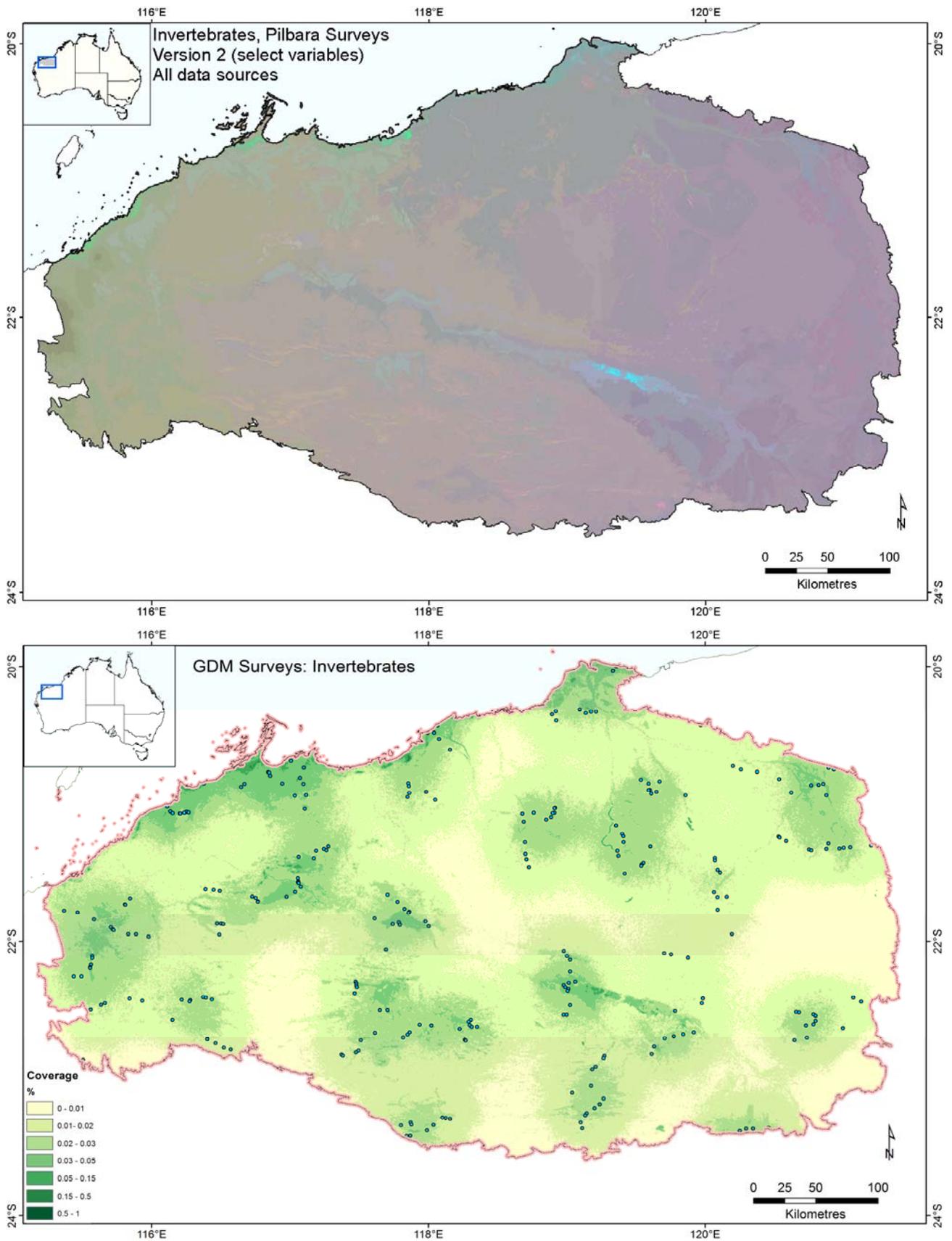
Figure 18. Classification of compositional turnover (upper) and estimated prediction uncertainty based on sampling coverage within GDM-scaled environmental space (lower) for invertebrates, shown for the Pilbara bioregion only.
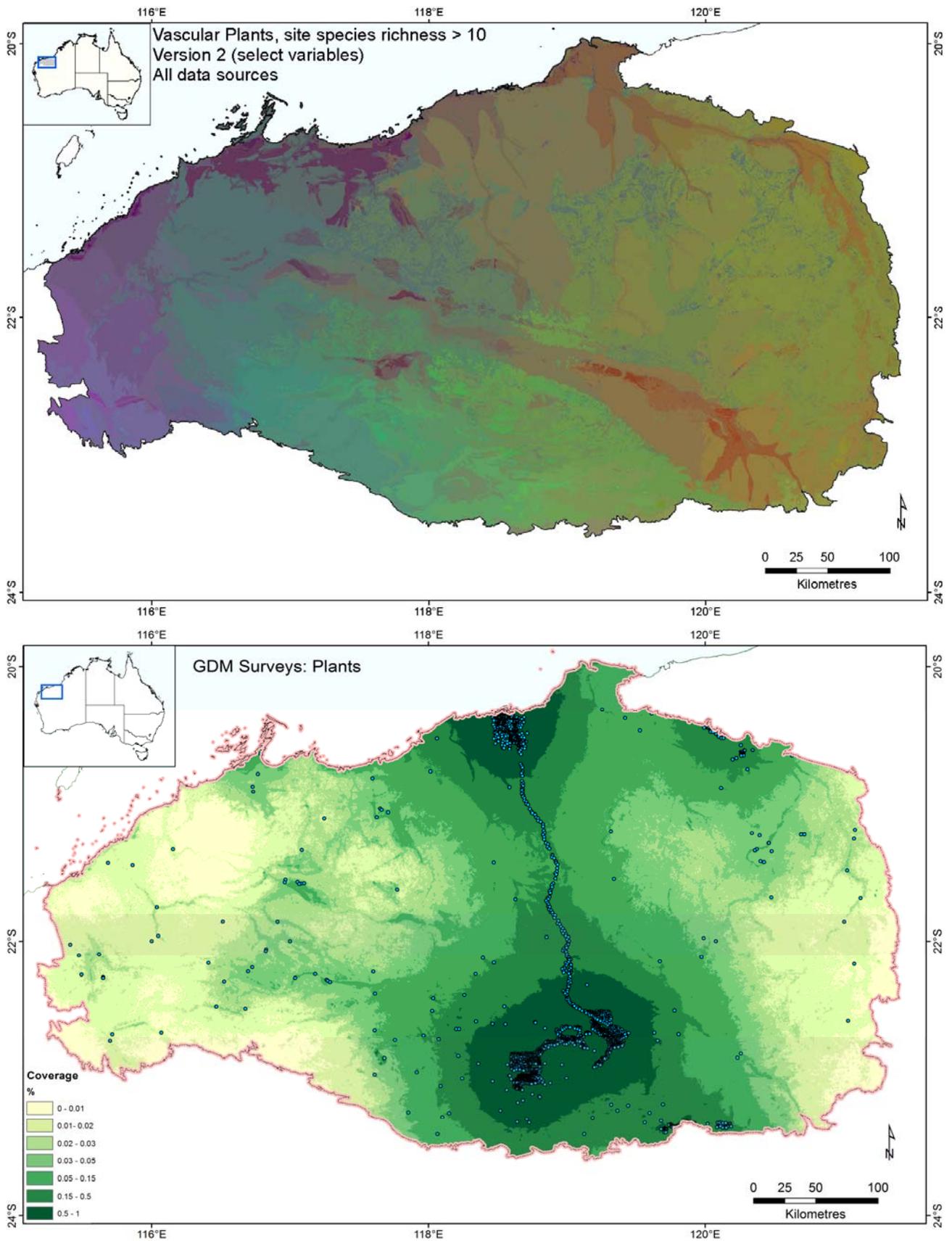
Figure 19. Classification of compositional turnover (upper) and estimated prediction uncertainty based on sampling coverage within GDM-scaled environmental space (lower) for vascular plants, shown for the Pilbara bioregion only.

3.1.4 DISCUSSION

The modelled patterns of compositional turnover are used in Section 5 to estimate "biodiversity significance" across the Pilbara bioregion. The associated estimates of uncertainty, derived from the sampling coverage help to identify those areas which are well supported by the model and other areas for which the environmental space is less well represented by survey locations. The maps depicting a classification of compositional turnover show areas of the Pilbara bioregion that are relatively distinct from each other in terms of the species that may be encountered.

The classification of predicted compositional turnover for birds suggests some distinct contrasts in assemblages across the Pilbara (Figure 16). This analysis is supported by relatively even sampling, drawing upon the numerous surveys compiled by Birds Australia (sourced via the Atlas of Living Australia), in addition to the Pilbara Biodiversity surveys (George *et al.* 2011) and consultancy data compiled by BHPBIO (Figure 8). This contrasts with the more sparse sampling for mammals, based largely on the Pilbara Biodiversity surveys (George *et al.* 2011) supplemented by other sources (Figure 7, Table 4). Vascular plant surveys however are largely drawn from BHPBIO sources, supplemented by the Atlas of Living Australia (Table 4) resulting in a strong spatial bias (Figure 11). The strong spatial bias in vascular plant surveys results in high confidence in predictive model outputs surrounding BHPBIO's mining tenements (Figure 19). While this will have influenced model fit, quite distinct assemblages are apparent across the Pilbara between the western coastal zone, north and inland that broadly align with the subregions (Figure 3), demonstrating the robustness of the GDM approach.

As requested by BHPBIO, we also tested the usefulness of the genus-level taxonomic rank for invertebrates in compositional turnover model, including additional survey data (Table 3). This test was considered necessary to account for the substantial additional data available at the genus rank, but lacking species level identifications (Table 3). Several models were examined – including all additional sources (DPaW, ALA and BHP), just including BHPBIO, or ALA, or DPaW data. Each model varied significantly and demonstrated sampling bias; reducing confidence in the applicability and usefulness of the model in subsequent biodiversity significance assessment. Therefore these models are not reported here.

The models of species richness patterns (detailed in Section 3.2, below) complement the compositional dissimilarity models (this section) for the estimation of "biodiversity significance" (detailed in Section 5).

3.2 Generalised additive modelling of species richness patterns

3.2.1 INTRODUCTION

Species richness patterns were modelled using generalised additive modelling (GAM), a statistical model developed by Hastie and Tibshirani (1990) that expands on generalised linear models by replacing the linear form with a sum of smooth functions. The smooth functions can be flexibly estimated using parametric or non-parametric methods and can take many forms. The model specifies an error distribution (such as a normal, or binomial distribution) and a link function relating the expected value of the distribution to the predictor variables with parameters specified by smooth functions such as splines.

Species richness data are counts of the aggregated number of species at a site (in this case defined by the 9-second, approximately 250m, grid cell) for a particular biological group. Count data type are usually analysed using the Poisson error term and log link function. However count data from species distributions at higher taxonomic rank (e.g., Order, Class) are classically zero-truncated (no zero values) because at least one species from the group is typically present; and absences are rarely observed systematically (taxon surveys are conducted where the taxon is expected to be found). This truncation requires modelling approaches that test or account for this type of data. The simplest approach is to analyse these data on a natural scale using the (log) Normal distribution. Other statistical models such as the truncated Poisson process or zero inflated Poisson process may also be suitable and some researchers advocate a negative-binomial error distribution (Barry & Welsh 2002; Leathwick *et al.* 2006; Potts & Elith 2006).

In this section we outline the approach taken to model species richness for the five biological groups: birds, mammals, reptiles, vascular plants and invertebrates (comprising ground-based beetles, spiders and ants), and provide some discussion of further ways to improve the modelling.

3.2.2 METHODS

As outlined in Section 2.4, comprehensive survey data are needed for the estimation of species richness. Suitable data sources were compiled from BHPBIO and DPaW. These data were aggregated within 9-second grid cells to provide the observations of species richness used in subsequent analysis. Because of uncertainty about the site-level sampling comprehensiveness of the biological survey data available from BHPBIO (Table 7) due to noted variation in survey method among consultants, initial investigations explored a minimum number of species per site (i.e., further truncating the count data) to account for potential (under) sampling variation in the case of birds, mammals, reptiles and vascular plants. For example, minimum richness per site of 2 or 3 species were explored for mammals and reptiles, and at least 6 species for vascular plants. This is consistent with the strategy adopted for the GDM models to account for under-sampling bias. We found this truncation had little effect on the richness model fit and so the fitted models presented here used all available richness data. A minimum number of species per site was not required with the invertebrate data which derive from a comprehensive survey using consistent field procedures (George *et al.* 2011) and at least 11 species were observed (Figure 10).

We drew upon a large number of candidate predictors for climate (27), regolith (35), landform (15), hydrologic and land cover (6) (detailed in Appendix B) to explicitly account for spatial dependence in the observed richness data. With such a large number of environmental variables, model parsimony, interpretability and over-fitting needs to be managed carefully. Exploratory models based on Random Forests (Breiman 2001), applied using the R package (Breiman *et al.* 2013), guided the choice of a sub group of variables to reduce the time and effort spent on subsequent fitted model development and selection. Random Forests are a data-driven process of variable selection that rank the importance of variables in a regression problem.

We developed fitted models for species richness using the generalized additive modelling (GAM) technique described in Wood (2006), as implemented in the R-package "mgcv" (Wood 2013). This approach uses a flexible regression framework that captures both linear and non-linear relationships between environmental variables and species richness. A stepwise GAM (`step.gam` function) was used to examine

all alternative models for the reduced set of variables determined from the Random Forests models. For each term in the process there is an ordered list of alternatives and the function traverses these in a greedy fashion, enabling a best model to be identified. In analysing the zero-truncated count data, we found that the Gaussian (Normal) Family with an Identity (Log) link function was the most parsimonious approach enabling a quick solution using well established R software packages.

Regression Trees were used to explore potential variable interactions. Qualitatively, the tree plot does a good job of capturing interactions between environmental variables. Putative interaction terms were tested in the GAM model and, if they improve the model, they were retained. Many quite different models demonstrate similar predictive ability (e.g., their % variation explained was consistent). While a higher deviance explained may be achieved by increasing the degrees of freedom, for example, the model becomes data driven and potentially less interpretable.

The fitted models were extrapolated using the spatial covariates to derive a spatial prediction and the standard error of prediction (described as "GAM uncertainty" in Section 3.2).

3.2.3 RESULTS

The range in observed species richness per site and the number of sites available for analysis varied between biological groups (Table 10, Figure 20). Low levels of richness relative to the median number of species observed at a site are apparent for birds, plants and possibly also, reptiles. The richness curve for invertebrates is indicative of a more comprehensively sampled biological group (Figure 20), the data having derived from systematic survey by McKenzie et al. (2009).

Table 10. Cardinal statistics for species richness observations for the five biological groups: mammals, birds, reptiles, invertebrates (ground-based spiders, beetles and ants) and vascular plants. The minimum and maximum values are based on data locations aggregated within 9-second grid cells.

BIOLOGICAL GROUP	MINIMUM	MAXIMUM	MEAN	MEDIAN	10TH PERCENTILE	90TH PERCENTILE
Mammals	1	9	2	1	4	1
Birds	1	62	14	14	29	1
Reptiles	1	28	8	8	15	1
Vascular plants	1	84	22	21	40	6
Invertebrates	14	94	41	39	58	27

Table 11. Summary of fitted richness models for the five biological groups: mammals, birds, reptiles, invertebrates (ground-based spiders, beetles and ants) and vascular plants (see details in Appendix D).

BIOLOGICAL GROUP	SAMPLE	% DEVIANCE EXPLAINED	R^2 ADJUSTED	GCV SCORE[1]	SCALE ESTIMATE	NUMBER OF PREDICTOR VARIABLES	MOST IMPORTANT PREDICTOR[2]
Mammals	1243	26.2	0.223	1.6594	1.5759	8	SLPFM300E
Birds	551	39.4	0.342	87.39	80.231	8	Interaction between DISTCOAST, EDISTFORST
Reptiles	558	38.0	0.336	22.587	21.055	9	Interaction between RAINIE, RAINXE
Invertebrates	403	29.1	0.271	62.492	60.673	4	Interaction between EDISTHYDRO, EDISTPEREN
Vascular plants	2492	21.8	0.197	148.21	144.3	9	PC2ME

1. GCV score: explained in Appendix D , page 95. 2. SLPFM300E - 300m focal median of percent slope; DISTCOAST - Euclidean distance to coast; EDISTFORST - Euclidean distance to woody vegetation; RAINIE – minimum monthly rainfall; RAINXE – maximum monthly rainfall ; EDISTHYDRO - Euclidean distance to water points; EDISTPEREN - Euclidean distance to perennial water bodies and drainage; PC2ME - Principal component 2 of spectra of surficial soils. Further details are given in Appendix B .

Exploratory Random Forests models were used to identify the variables with the most predictive power for the species richness data and, following a rigorous process of model selection including tests for interaction terms without adversely impacting parsimony, a set of fitted models were derived (Appendix D , Table 11). The richness models for birds and reptiles explained the most deviance and the vascular plant model, with the greatest degree of sampling bias, explained the least. Examination of the fitted models suggested the (log) Normal distribution assumption was reasonable in all cases except mammals, and therefore the (log) Normal model for mammals is not statistically valid.
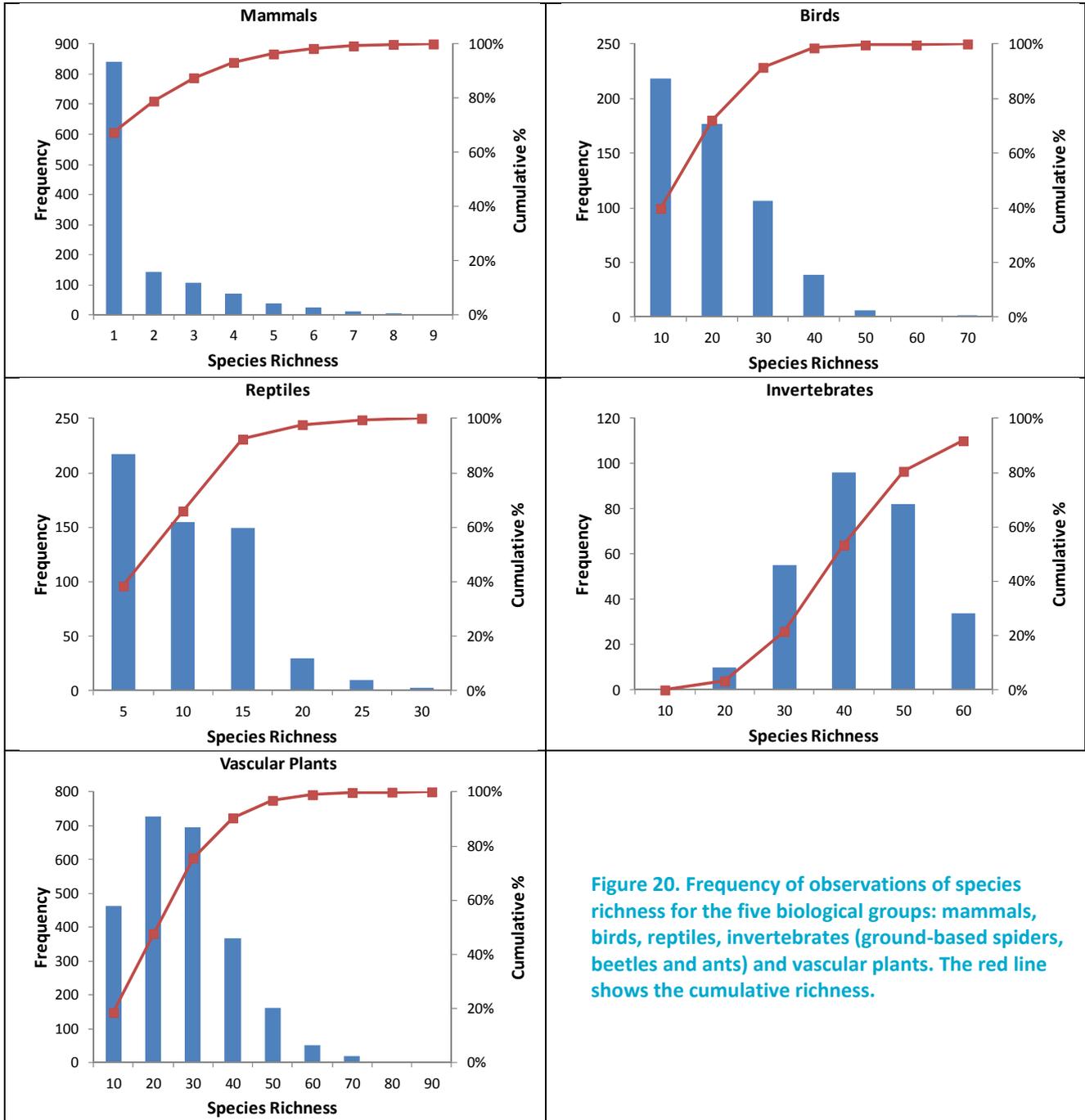
Figure 20. Frequency of observations of species richness for the five biological groups: mammals, birds, reptiles, invertebrates (ground-based spiders, beetles and ants) and vascular plants. The red line shows the cumulative richness.

Maps of the spatial predictions and the standard error of prediction for each group are presented in Figure 21 to Figure 25, for their extents only within the Pilbara bioregion. The prediction maps are shown with an overlay of the standard error normalised within the range of the maximum predicted value to enable comparison of general uncertainty between models. The standard error maps are shown in context with the location of data used to model richness. Areas of low standard error indicate less certainty and higher standard errors indicate greater uncertainty in the spatial prediction, often associated with increasing extrapolation beyond the data range or weaknesses in the fitted model (see also individual predictor

functions in Appendix D). For example, the lower map in Figure 21 shows a region of high standard error of predicted species richness for mammals associated with the Fortescue Marsh. This region is masked due to model uncertainty in the upper map in Figure 21 showing the predictions of mammal species richness.
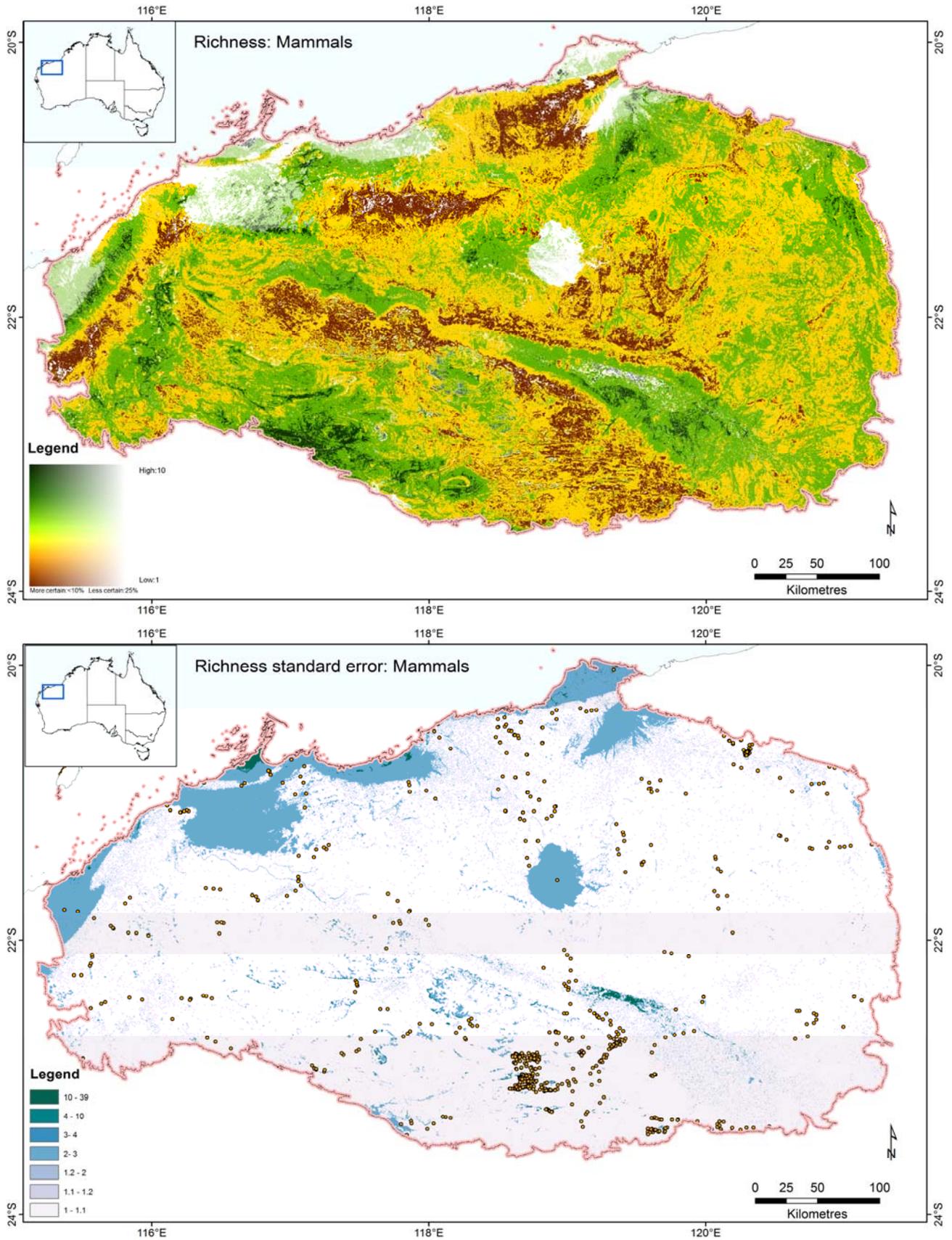
Figure 21. Spatial prediction of richness (upper) and standard error of prediction (lower) for mammals.
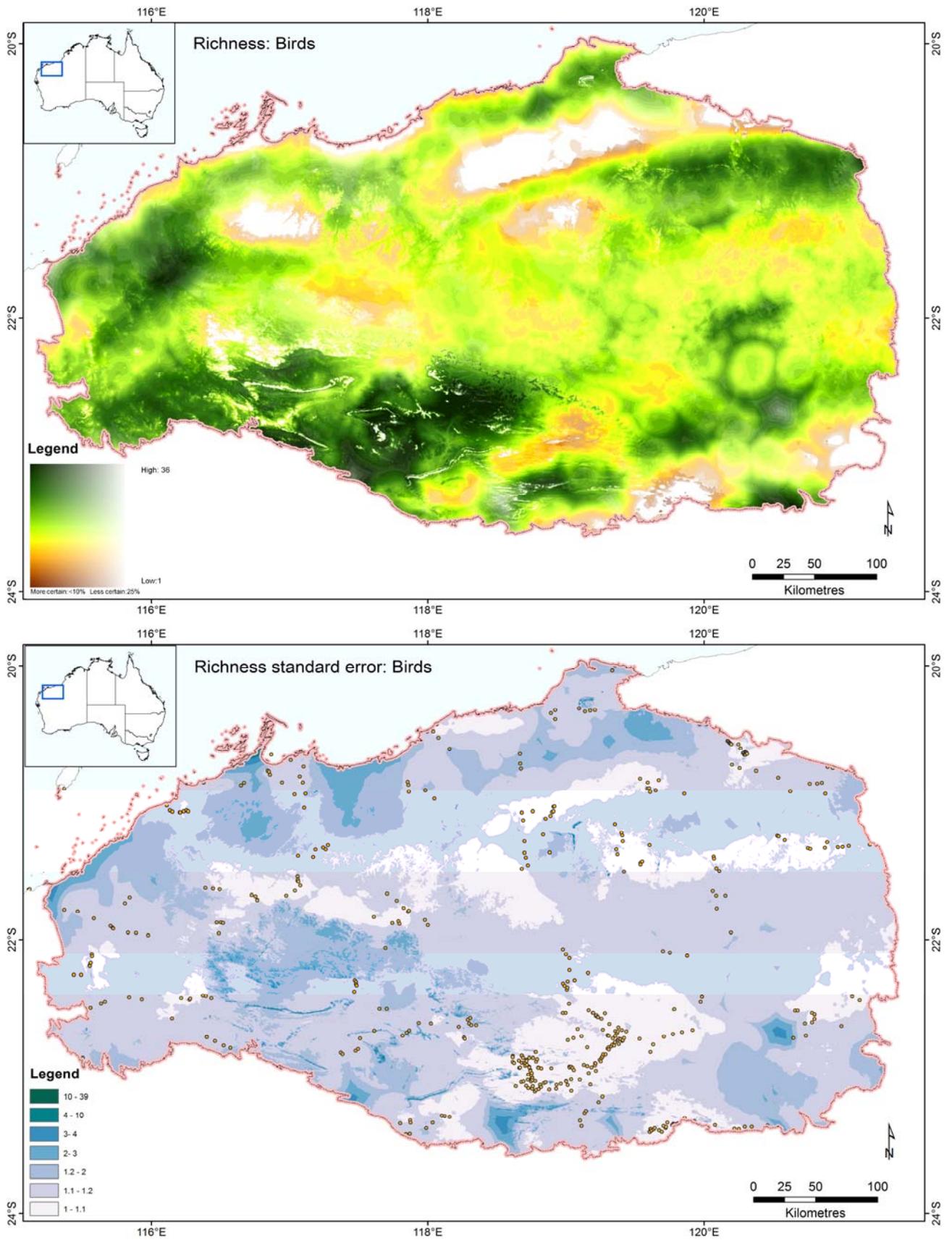
Figure 22. Spatial prediction of richness (upper) and standard error of prediction (lower) for birds.
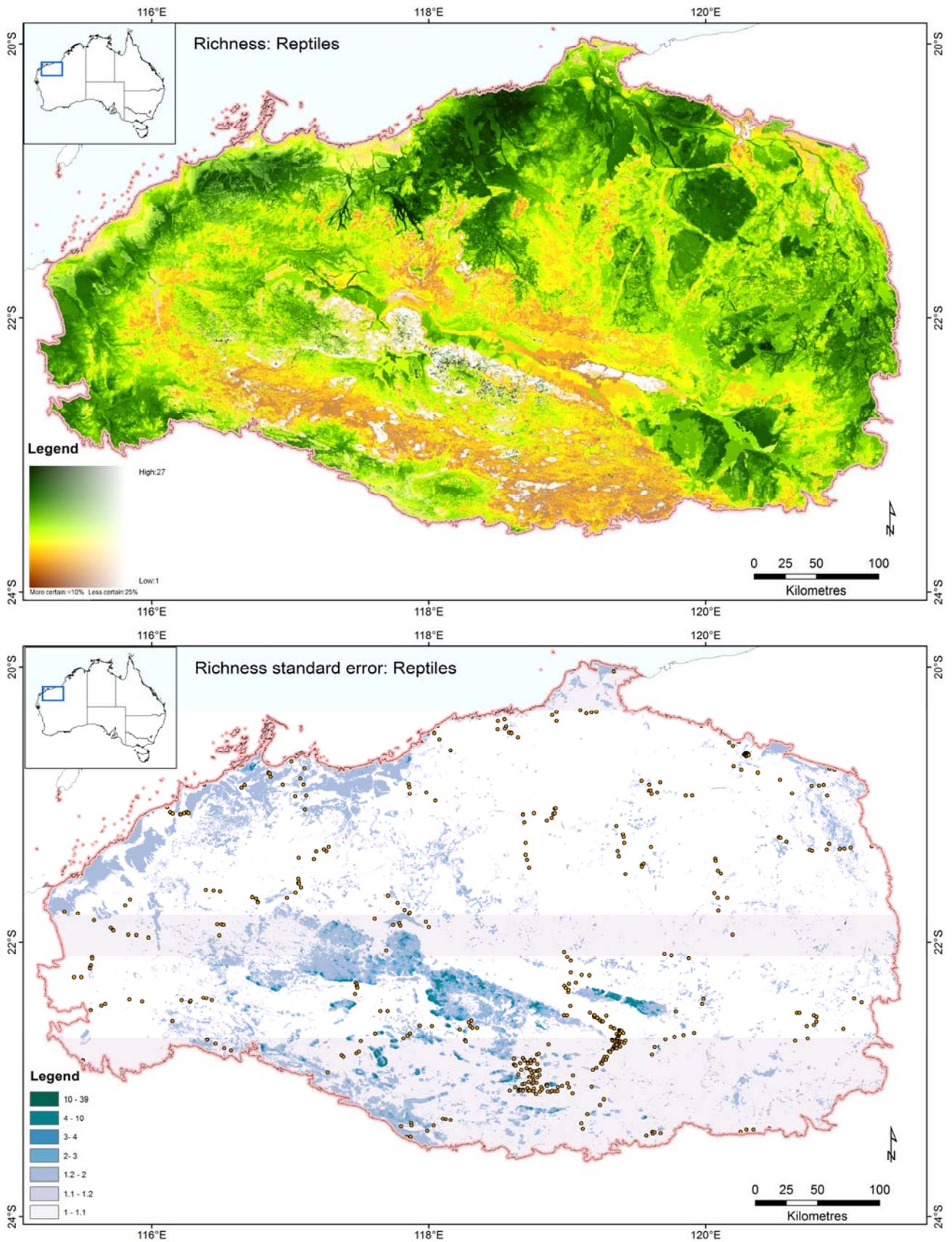
Figure 23. Spatial prediction of richness (upper) and standard error of prediction (lower) for reptiles.
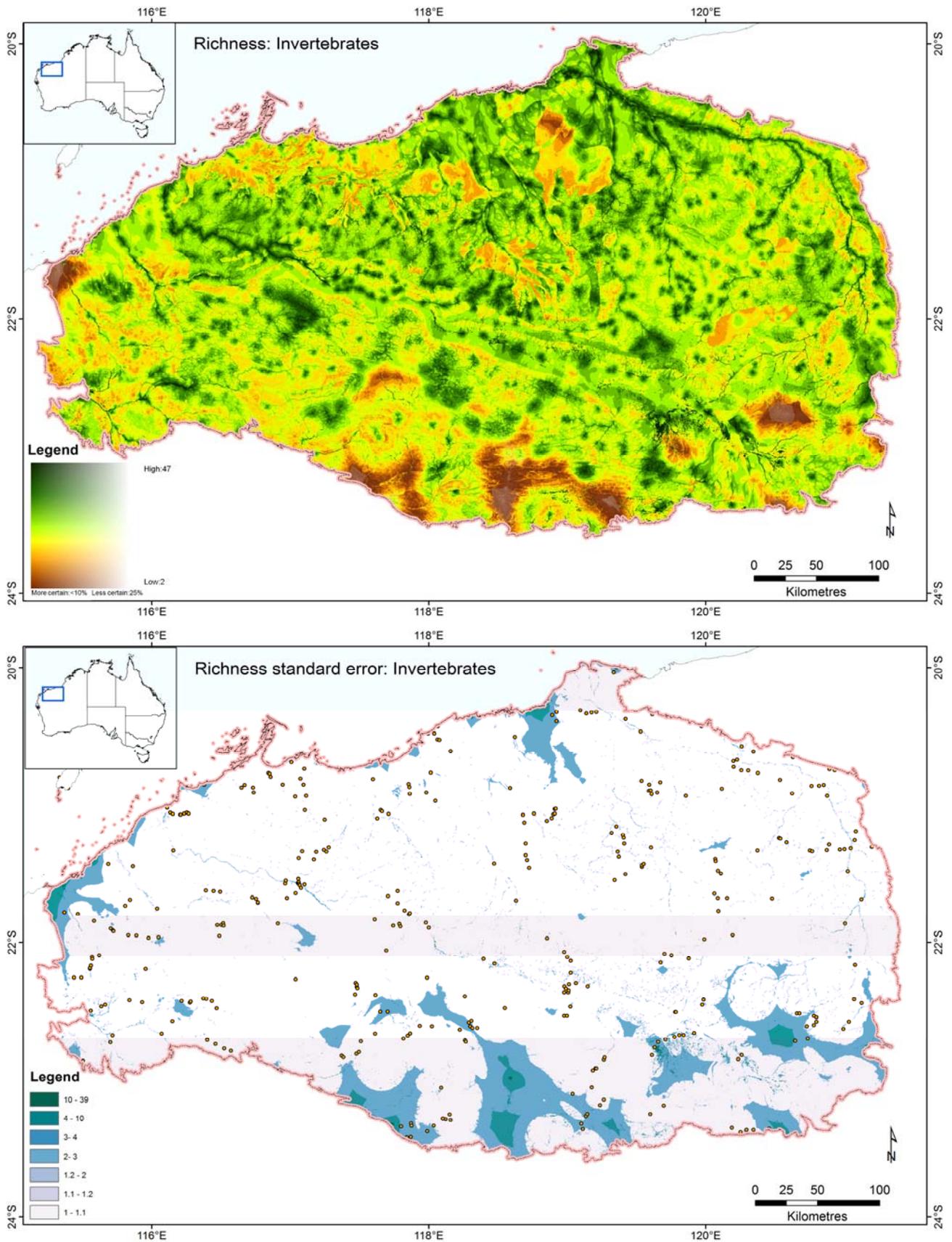
Figure 24. Spatial prediction of richness (upper) and standard error of prediction (lower) for invertebrates.
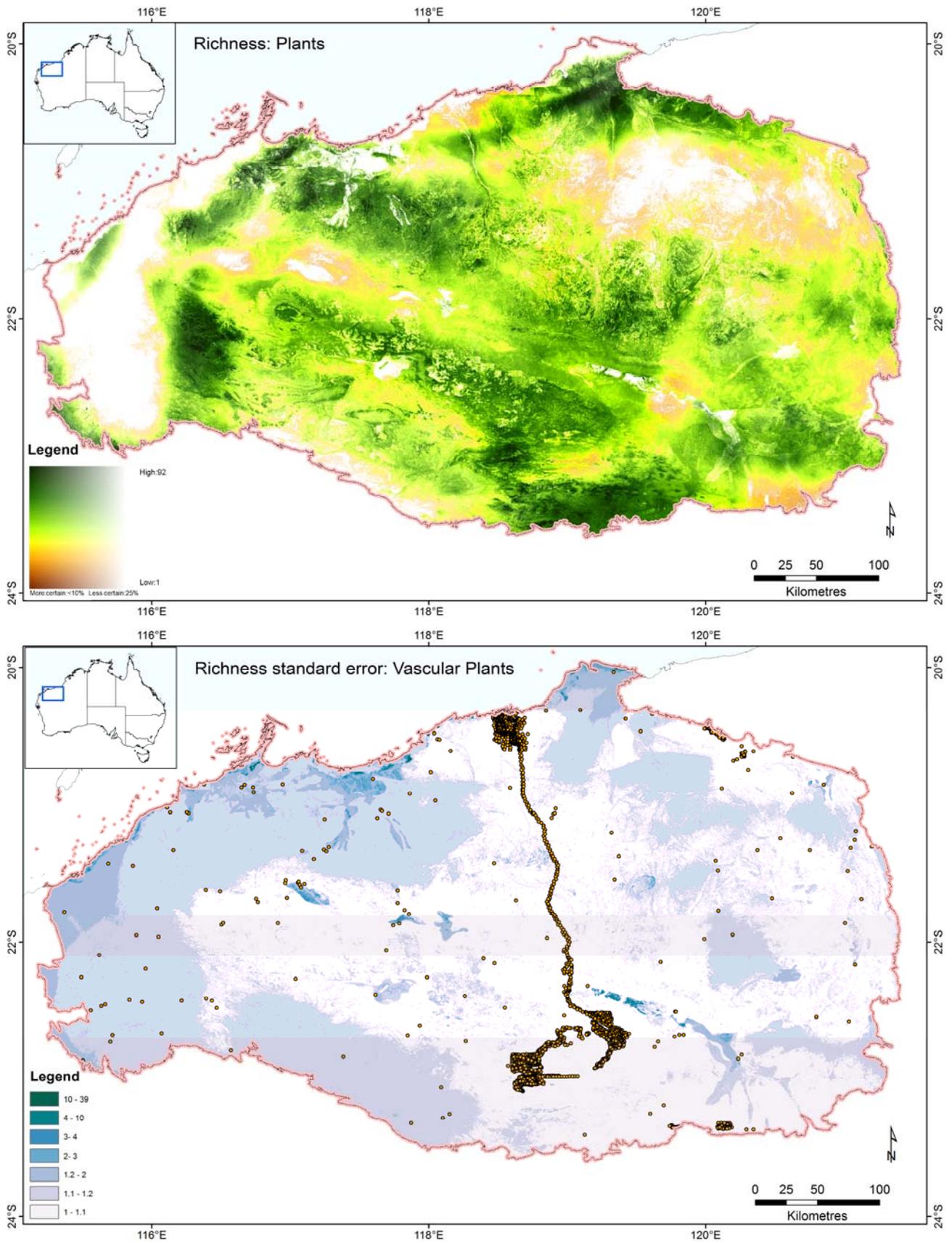
Figure 25. Spatial prediction of richness (upper) and standard error of prediction (lower) for vascular plants.

3.2.4 DISCUSSION

Our modelling of species richness has highlighted why there is continuing scientific debate about the best way to analyse and spatially extrapolate such data (Barry & Welsh 2002; Leathwick *et al.* 2006; Potts & Elith 2006). While species richness has traditionally been equated with count data, the data we compiled are better described as zero truncated presence-only data for which the count starts at 1 species or more. The observation of species richness depends on the prevalence of species within a group, the extent of the study region, and the comprehensiveness and representativeness of sampling. The ideal data for this analysis derive from systematic and comprehensive surveys that are representative of all environments across the study region of interest and surrounding areas. However, the data we compiled necessarily derive from different sources that vary in either sampling method or comprehensiveness; generally without environmental stratification (except the Pilbara biodiversity survey data). In preliminary models we explored further truncating the richness values to reflect ecological expectations of a minimum number of species observed in a survey; similar to the thresholds used in the compositional turnover models (Table 9); but found this has little effect on the richness model fit and substantially reduced the data available to the analysis. A full assessment of this data variation, and whether richness thresholds should be used, requires a closer examination of the survey methodology associated with each source and the development of predictor covariates describing how each survey differed so these can be more effectively combined in a model.

For these data, the statistical modelling assumptions of Poisson regression traditionally used for count data were found to be invalid. Poisson regression assumes a Poisson process in which the count starts at 0. Therefore all models were developed using (log) Normal regression. The statistical assumptions of the (log) Normal regression applied to the species richness data were found to be reasonable in the case of birds, reptiles, invertebrates and vascular plants (see model checking plots in Appendix D , e.g., Figure 53). The QQ plot is relatively close to a straight line, the variance is approximately constant as the mean increases, and the histogram of residuals appears consistent with normality (although slightly skewed). In the case of mammals, however, the qq plot departs from the straight line and there is asymmetry in the histogram of residuals (Figure 51). Therefore the log (Normal) statistical distribution assumption for mammals is not statistically valid. The mammal count data are dominated by single species and there are few levels (values range from 1 to 9) (Figure 20). More work is required for the mammal model to be used with confidence in subsequent analyses. The arbitrary importance of the landform predictor in the mammal richness model representing neighbourhood topographic variation due to slope (SLOPEFM300, Table 11), which lacks ecological credibility, is also indicative of the low confidence in this model.

We also have some concerns about the validity of the statistical assumptions for the vascular plant richness model (Appendix D , Figure 59), although the departure from Normality is much less extreme than for mammals. Spatially, the richness data for plants are highly reliant on the survey data provided by BHPBIO (Figure 12). Supplementary data from the Pilbara biological surveys are only associated with riparian flora surveys (Pinder *et al.* 2009), which only weakly contribute to the fit of the model with large areas of uncertainty remaining as evident in the richness standard error (Figure 25). Terrestrial flora survey data (George *et al.* 2011) were not available for this analysis.

The Random Forests models provide a good baseline indication of the amount of explanatory power that is possible for the richness data given the spatial and environmental predictor data (Appendix D , e.g., Table 21). Considering also the regression models (Table 11), this accounts for around 30-40% of the variation. Richness is therefore not highly predictable but still explains a reasonable amount of the deviance. While deviance explained is not a primary basis for choosing between alternative models or among the models for the different biological groups; how reasonable the statistical assumptions are is a key consideration, as discussed above.

The spatially estimated standard error for each model (e.g., Figure 21) provides an indication of prediction credibility but there is no guarantee these values make sense in practice. For example, if we model the entire region based on data from area A, then predict area B using this model, we can expect the predicted value to have a high standard deviation (probably extrapolated). On the other hand if we use the same model to predict a location that is also in area A then we would expect the predicted value to have a much

smaller standard deviation (probably interpolated). In the few cases where the model predicts negative values (<<0.01%) these were replaced with zero values for mapping. While in theory the absence of zero values in the training data does not imply zero predicted values are less credible, zero values for species richness at the high taxonomic levels (birds, mammals, etc) of this analysis are expected to be rare events and potentially spurious. Some locations may in fact be naturally devoid of species within a particular group, but the data used here are based on presence-only counts. Each model predicts species richness based on the available data and interpolates to fill gaps between known points; and with increasing extrapolation, the standard error increases.

In our mapped presentation of the predicted values (e.g., Figure 21) we include a scaled representation of the standard error as an overlay. This scaled standard error enables consistent comparison among richness models for the different biological groups to highlight areas where the prediction is reasonably interpolated compared with areas of likely extrapolation. The "uncertainty" overlay is scaled in the range 0 (more certain) to 1 (less certain); regions where the standard error is greater than the predicted value were set to 1 and excluded (white without transparency). Similarly, zero predicted values were treated as areas of high uncertainty and also depicted as white without transparency.

These results are based on our analysis so far, and with more investigation there is scope for improvement. We suggest the following extensions and refinements to build credibility in the richness models for use in biodiversity significance assessments:

1. The data have been analysed on a natural scale using (log) Normal distribution, and could be considered more correctly as count data, requiring that the zero-truncated nature of these distributions be addressed using zero-inflated Poisson process methods (e.g., Zuur *et al.* 2012) and potentially also make use of relevant functions in 'VGAM' R package (Yee 2013).
2. Some of the spatial associations of the data are handled implicitly through the correlation with the environmental predictor variables (covariates). However, the spatial dependence in the observed species richness could also be modelled explicitly using statistical auto-correlation methods. A starting point would be to examine whether there is spatial structure remaining in the residuals through a variogram or covariance analysis. If there is, potential exists to use both the covariates and spatial proximity among observation locations in the model.
3. The spatial sampling strategies used play an important role in what is observed and their impact needs to be considered more carefully. This is arguably even more important for presence only data. In some cases, it makes sense to reweight the sampling so that more random samples are drawn. Additional value may be derived also by considering spatial sampling variation using covariates derived from a closer examination of each survey's methodology (e.g., time of year, experience of surveyor, sampling duration or intensity, sampling methodology, etc).
4. There are a large number of environmental covariates and potential interactions. While this was considered in the models developed and maps created, there is no doubt greater effort could be directed to identifying the most appropriate model by considering a wider range of potentially useful variables.
5. It may be worth also considering developing spatial predictions for some of the more prevalent individual species using the same set of candidate environmental variables. This process could augment the sample by filling gaps in the training dataset, if the resulting models themselves were reasonably valid. With careful consideration of the assumptions and credibility of this process, new estimated data points could be defined within the range of a discrete set of known data points. Ideally, however, new observations in unsampled environments would be used to augment the data as a basis for modelling.

4 Habitat condition mapping

4.1 Introduction

A spatially explicit expression of biodiversity habitat condition is an important component of biodiversity significance assessment. Our definition of habitat condition is designed to work flexibly with a biodiversity assessment and modelling framework for estimating the overall regional retention of native biodiversity (Zerger *et al.* 2013).

The "condition" of land however means very different things to different people. To a crop grower, condition may relate to the land's capacity to support crops or to a pastoralist, the abundance and nutrient value of the sward. From a remote-sensing perspective, condition is often viewed as something which changes between satellite images, such that, for example, fire scars or flood deposition will be detected and will reduce the condition score for a given pixel. From the perspective of biodiversity assessment, however, condition describes an area's capacity to support functioning ecological communities which are appropriate to its environmental state. As such it may be possible for a place to undergo a dramatic shift in vegetation abundance and/or species composition and consequently its remotely sensed signature. From the perspective of habitat condition, therefore, this represents natural variation; that is, dynamic or temporally-related variation that is often represented by a spatial mosaic at any point in time. In this definition, condition is measured in relation to natural environmental variation rather than in relation to an abstract ideal.

Quantifying habitat condition at any scale is a very demanding task and there is a large body of scientific literature illustrating the inherent complexity underlying disturbance ecology (Lindenmayer *et al.* 2003). In the past 200 years most Australian ecosystems have been dramatically altered directly and indirectly by changes in land use and management, primarily driven by agriculture, leading to unprecedented declines across all vertebrate taxa (Ford *et al.* 2001; Johnson 2006; Kutt & Fisher 2011). Northern Australia has not been immune to changes although the rate of change has been slower and less conspicuous and therefore more difficult to record (Lewis 2002). It is often purported that Australian tropical savannas and rangelands are largely intact and, relative to other global rangelands, this is true (Woinarski *et al.* 2007). However, in the past century significant disruptions to indigenous burning regimes (Russell-Smith *et al.* 2009) coupled with an intensification of the pastoral industry (Dyer 1997; Fisher *et al.* 2004; Woinarski *et al.* 2011) has in many cases led to significant structural and floristic vegetation changes via tree clearing, the proliferation of exotic pests (both plants and animals) and degradation of wetlands and rivers.

The coincident changes in fire management and intensified herbivory by domestic stock have had significant cumulative impacts on grass and other ground cover species both by selective grazing and too frequent burning which can eliminate fire prone species such as *Triodia* (Crowley & Garnett 1998), or alter tree-grass balance by promoting the dominance of trees and shrubs over grasses (Crowley & Garnett 1998; Crowley & Garnett 2001). Species are impacted in different ways by disturbance depending on their life history traits. For example, declines in a variety of grasses has been linked with seasonal resource bottle necks, where important seed resources are not available at critical times of the year, leading to a reduction in suitable habitat for gravinorous birds and mammals (Franklin 1999; Woinarski *et al.* 2011). Other vertebrate declines have been strongly linked to structural changes in habitat, such as loss of cover from predators and removal of nesting sites, or a disruption of other food sources, such as a reduction in invertebrate abundance (Perry *et al.* 2011a). There is little knowledge or at least a high level of uncertainty as to the impact of landscape change on reptiles across the rangelands (Woinarski *et al.* 2004; Mott *et al.* 2010). However, shifts in vegetation structure have been shown to alter the thermal environment for heliothermic (body temperature controlled by heat from the sun) species, leading to increased susceptibility of some species to changes in climate and weather patterns (James 2003; Huey *et al.* 2009; Sinervo *et al.* 2010).

In northern Australia, pasture condition is often confused with habitat condition under the misplaced assumption that what is good for domestic live stock will favour a healthy biodiversity (Fisher & Kutt 2007). However, by definition 'habitat' is species-specific (Hall *et al.* 1997) and to adequately assess the condition of habitat in any given area there is an underlying assumption that the thresholds for change under a disturbance regime and the associated threatening processes are known for each species and accounted for in the analysis. It is also assumed that the available environmental variables used in the analysis represent the primary limiting factors or are adequate proxies for all the species in a region. Further compounding the complexity is that, what is considered to be poor condition for one species, may be highly favourable for another (James 2003).

To account for some of these assumptions, a landscape-scale condition assessment needs to be nested within land types representing a set of relatively homogenous environments encompassing the regional variation. Land types essentially classify continuous features of the landscape into similar units, such that species that inhabit these areas are likely to have similar physiological limits and responses to variation in the stochastic and anthropogenic drivers that ultimately influence which species can be sustained. This is somewhat analogous to the community level analysis of species compositional turnover described in Section 3.1, where the environmental limits of biological communities are used to inform the scaling and categorisation of environments (e.g., upper map in Figure 19). A key objective of the landscape-scale condition assessment, however, is to separate the effects of natural disturbance regimes on vegetation and habitat dynamics from the effects of anthropogenic drivers of habitat modification and the cumulative processes of degradation. In practice however, these processes typically interact, resulting in gradual or episodic degradation.

4.1.1 CONTEMPORARY CONDITION METRICS

Regional condition assessment has long been the responsibility of government departments responsible for agriculture. As such, many of the indices and methods developed have been tailored for measuring pasture condition and tend to quantify the utility of land types for sustaining domestic stock. Various methods for measuring condition are enacted across the vast rangelands of northern Australia, a well published example of this is the ABCD land condition assessment (Chilcott *et al.* 2003). In the Pilbara region there have been various mechanistic assessments of condition over time including an inventory and condition assessment (Van Vreeswyk *et al.* 2004b), regional biodiversity assessments (George *et al.* 2011) and many smaller location specific assessments undertaken by consultants (BHPBIO, unpublished data). The most wide spread and temporally consistent monitoring in this region is the Western Australian Rangelands Monitoring System (WARMS) including over 600 permanent monitoring sites in the region (Watson *et al.* 2007a). The sampling protocol and subsequent placement of permanent plots was specifically designed for pasture monitoring and therefore the sites are not representative of the region but rather of a sample of production areas. In addition, these data are held in confidence by the Department of Agriculture and Fisheries of Western Australia (DAFWA) and are not released without the explicit permission of each land holder. While access to these data is feasible, the process, cost and time required for acquisition precluded their use in this project.

A broad assessment of pasture condition throughout the Pilbara has also been undertaken using what is known as a traverse (Van Vreeswyk *et al.* 2004a). A traverse is a rapid assessment of condition made at 1km intervals along roads and tracks within the region. 12,450 individual assessments of condition where conducted in the Pilbara region using the traverse method. Results from this rapid survey from 2003/04 suggested 77% of vegetation was in good or very good condition, 11% fair condition and 12% in poor or very poor condition. Around 0.2% of the region was listed as in critical condition (referred to as severely degraded and eroded areas), areas lacking any perennial vegetation and with no grazing or ecological value. Similar to the department of agriculture's condition assessment (Watson *et al.* 2007a), the traverse method measures pasture condition only and cannot be equated with an assessment of biodiversity habitat condition. However, some of the characteristics of areas recorded as being in poor condition may have relevance to a measure of habitat quality for biodiversity.

Historically there was a view that condition change was gradual and linearly related to manipulated stocking rates. That is, a reduction in stocking rates or de-stocking would result in pasture condition improvement. However, consistent rangeland monitoring over time (Watson *et al.* 2007b) has demonstrated that major changes occur during climatic extremes and subsequent impacts on pasture composition can persist well beyond the stochastic event if poor land management continues. For example, there is strong evidence that transformative change to ground cover and vegetation composition occurred in the Pilbara region following the intensification of stock (mainly sheep) during a decade of above average rainfall in the 1930's followed by extreme and persistent drought (Figure 26).
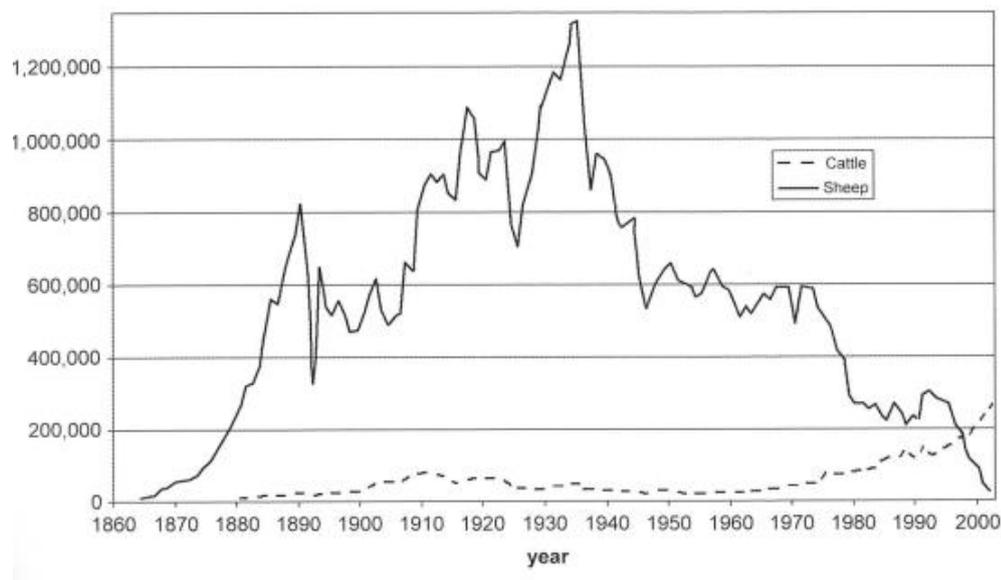
Figure 26. Stock numbers in the Pilbara region 1860 – 2000 (Van Vreeswyk *et al.* 2004b) illustrating a drop in sheep numbers from the 1980's and the continuing expansion of the cattle industry.

The sheep industry did not have the adaptive or predictive capacity to dramatically reduce stock in response to the rapid change in land condition, resulting in the dramatic landscape degradation (erosion, persistent bare ground and changes to vegetation composition) that still continues today. Another key consideration is that following transformative change landscapes can shift into a new stable state of depletion that may not respond to altered land management (Kutt *et al.* 2012). That is, the pathway to degradation and subsequent decline of biodiversity is linear but recovery from that degradation does not follow the inverse pathway. This demonstrates the need for tools that can dynamically predict changes in land and habitat condition in the context of biodiversity. This will require condition metrics that consider multiple threatening processes at multiple time scales and specifically for this region the cumulative impact of infrastructure development following the expansion of mining.

By completing a landscape scale assessment of ground cover we can identify the areas in the landscape that are responding atypically to stochastic events (wet years and dry years). To effectively do this we need to quantify what is a normal response to weather in vegetation cover (bench mark condition) for each land type (Bastin *et al.* 2012). The bench mark condition, once defined, can be used to dynamically identify areas that are above or below the normal trajectory of change.

There are a number of challenges involved in generating a useful habitat condition layer for biodiversity management planning and assessing status and trends. These include the availability of suitable data sources describing;

1. Extent and density of environmental weeds.
2. Distribution and densities of detrimental non-native species, including domestic livestock.
3. Knowledge of the impact of the timing and intensity of fire on species.
4. Adequately scaled vectors or rasters for representative land types.
5. Temporal ground cover indices at an appropriate scale.
6. Accurate regional monthly weather grids.

7. Historical information on disturbance.
8. Current distribution of land use and intensity of uses.

In this section we refer to a preliminary analysis of vegetation cover change across the Pilbara region that we conducted but which requires further work to be useful in decision making (Appendix E), and present the results of an interim analysis of habitat condition based on best-available regional data to demonstrate how such data informs the analysis of biodiversity significance (Section 5).

4.2 Method

Two methods of condition assessment were undertaken. In Appendix E we present a satellite-based assessment of vegetation cover change (gains and losses), provide examples of its utility and outline the requirements for refining these data in the future. Without contextual management data to use in calibrating the change analysis with ecosystem condition, these outputs were not ready to apply in the subsequent biodiversity significance assessment. We therefore undertook a rapid assessment of habitat condition (outlined in Figure 27) at a broader resolution using the best-available, accessible datasets for land use, tenure and infrastructure as indicators of the potential for human modification of ecosystem function. We developed and applied an additive scoring method to infer habitat condition which was used as a demonstration in the subsequent analysis of biodiversity significance (Section 5). This section describes the data inputs and how they were combined to generate a condition index.
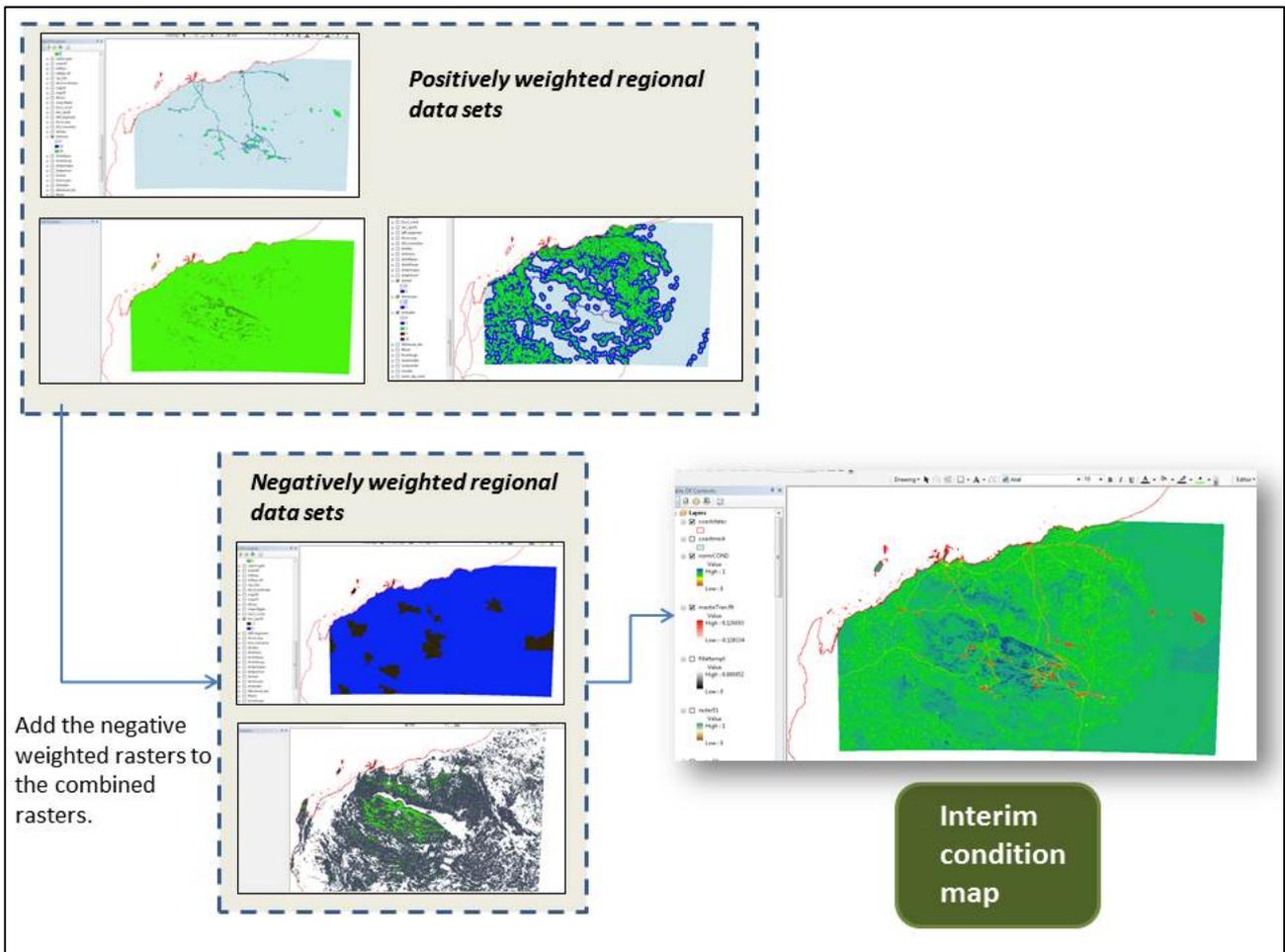
Figure 27. Schematic showing steps in the rapid assessment of habitat condition using best-available datasets for land use, tenure and infrastructure. The diagram considers the scores in one simulated pixel and sums those values. Across the Pilbara, the final scores were range-standardised between 0 and 1, and inverted so that 0 indicates completely removed and 1 indicates intrinsic condition.

Compiling historical data representing disturbance impacts

Various disturbance processes that threaten the persistence of biodiversity have been identified in the Pilbara region. These include expanding weed infestations, density of feral animals, inappropriate fire regimes, rapidly expanding mining infrastructure and over stocking of domestic sheep or cattle (George *et al.* 2011). However, with the exception of grazing and mining, there were no other appropriate mapped indicators of other threats to biodiversity-habitat condition at a landscape scale.

One of the primary sources of landscape change in northern Australia has been the proliferation and intensification of domestic livestock, primarily sheep and cattle. In some cases the habitat disturbance impacts have been subtle, occurred incrementally and gone largely unnoticed (Lewis 2002). In other cases, the changes have been rapid and dramatic such as broad scale tree removal and replacement with buffel grass in central Queensland (Kutt & Fisher 2011; Kutt & Kemp 2011). There is also evidence of stability in community composition over decades that may represent a depleted stable state. This followed the broad scale landscape change and associated species declines that occurred over many years before systematic data were collected (Perry *et al.* 2011b). Therefore, when measuring habitat condition it is important to account for the historical context of disturbance which may help to describe differences in species composition within otherwise similar ecosystems.

To account for sustained grazing pressure, historic station maps were acquired (Table 12). Station boundaries from the 1950's were geo-referenced against contemporary tenure maps using boundaries that haven't changed since the 1950's (Figure 28) using the geo-referencing tool in ArcMap10 (ESRI 2011). Water points and station dams were geo-referenced and saved out as a point file to create an historic map of watering points for the Pilbara region. The Euclidean Distance tool (ESRI 2011) was used to produce a raster map with each grid cell representing the linear distance to the closest historical water point. This value was derived as a proxy for grazing pressure with the assumption that areas close to a watering point are more likely to have endured sustained grazing intensity (Foran *et al.* 1982; Bastin *et al.* 1993). Pringle and Landsberg (2004) demonstrated a non-linear relationship between proximity to water and track density (a proxy for grazing pressure) where the greatest pressure was within 500m of a water point dropping by about 75% in 1km and then flattening out between 4km and 10km. We use these data to categorise weighted potential grazing pressure with decreasing distance to water (Table 13).

To accurately estimate the condition of the Pilbara region it is important to account for threatening processes at a variety of spatial and temporal scales. Biodiversity responds to disturbance processes resulting from the combined effects of land use and climatic extremes coupled with the various cumulative impacts mediated by the contemporary natural resource users of the region.

Small mammal abundance has been correlated with higher grass cover present inside conserved lands compared with grazing tenure in Australian rangelands (Kutt & Gordon 2012). However, there are several opposing views on the efficacy of national parks as conservation entities (Bowman 2012; Flannery 2012) and legislative protection of reserved land differs between states. Additionally, it is common practice for reserves and pastoral properties to use working boundaries that increase the efficiency and cost effectiveness of building and managing fence lines. These working boundaries can be significantly different to the official cadastral boundaries. Context is required to accurately measure the impact of reserved tenure on biodiversity. The authors acknowledge that management regimes outside of national parks are extremely variable and therefore the relationship will not be consistently positive or negative. Here, it is advantageous to combine tenure maps with linear models of cover change to demonstrate clear fence line effects.

We acknowledge that the resultant output has many underlying assumptions and does not account for variance in stocking levels, current water points, and recovery from previous grazing pressure, fence line affects and fires. However, this represents a starting point for refinement of condition layers. Furthermore the derived values that could be ascribed to grazing and mining leading to poor condition are untested for their impact on biodiversity and therefore remain an indicator of potential impact rather than a precise statement of habitat quality.

Quantifying the cumulative potential impact of mining on biodiversity requires measuring impact at a macro scale. To effectively analyse the impacts at this scale it is important to acknowledge the presence of

highly visible impacts such as pits and major mining infrastructure (Figure 62) as well as the cumulative impacts of roads, survey grids, plant, and rail that are not as visible but collectively can cover extensive areas (Figure 29 and Figure 30).

Table 12. Description and source of historical station maps for the Pilbara region (courtesy Brian Goodchild, Western Australia Department of Land Administration).

YEAR	SCALE	SHEET	DESCRIPTION
1948	10 mile 1:633,600	De Grey (Sheet 14)	Station boundaries and internal infrastructure
1949	10 mile 1:633,600	Nabberu (Sheet 11)	Station boundaries and internal infrastructure
1951	10 mile 1:633,600	Carnarvon (Sheet 10)	Station boundaries and internal infrastructure
1960	10 mile 1:633,600	Onslow (Sheet 13)	Station boundaries and internal infrastructure

Figure 28. Example of a geo-referenced station infrastructure maps from 1950. Small black dots represent water points and rectangles represent farm dams.

Figure 29. An example of mining infrastructure expressed at a coarse scale. At this scale the impact of mining appears localised.

Figure 30. When zoomed in and including the extent of mining infrastructure; such as roads, plant and rail; the extent of local disturbance is more apparent.

The intensification of infrastructure development often occurs incrementally and relatively unnoticed when each project is viewed in isolation from the cumulative effects of the others. Data on mining infrastructure was provided by BHPBIO for their tenements. A regional map of mining infrastructure for all of the mining companies was not available. To approximate the overall footprint of mining across the Pilbara (beyond BHPBIO), a mining tenement map (Geoscience Australia 2013a) was used to identify all areas with "live" mining tenement. Only those areas listed as having been surveyed were included in the analysis of mining impacts. In addition, an operating mines layer (last updated October 2012) was downloaded from the

Australian mines atlas website (Geoscience Australia 2013a). The operating mines layer and the filtered tenement map were ground-truthed using high resolution imagery in Google Earth (Google Inc. 2013). Polygons were drawn around areas with clearly visible infrastructure (i.e., grid lines, pits, buildings, extensive roads and tracks). Major rail infrastructure was sourced from the Australian mines atlas mapping portal (Geoscience Australia 2013a) and geo-referenced against the Western Australian tenure maps using ArcMap10 (ESRI 2011). The BHPBIO data were merged with the derived mining infrastructure layer using the merge tool in Arcmap10 (Figure 31). The Euclidean Distance tool in ArcMap10 was used to develop a raster grid with each pixel representing the distance in decimal degrees from mining infrastructure (9-second resolution). Aside from the fine scale data provided by BHPBIO, these data only approximate other industry mining infrastructure in the region and exclude associated mining roads and tracks (for which spatial data were not available and their derivation from imagery was beyond the scope of this project). Future analysis could be far more robust if regional data were available on infrastructure from all companies consistent with the data provided by BHPBIO. The distance to mining infrastructure was classified into two distance categories (500m, 1km) and weightings were applied to reflect relative ecosystem disturbance (Table 13).

Compiling historical data representing disturbance impacts

The distance to water and mining infrastructure were combined with classified and weighted values (Table 13) for distance to roads of different categories, rail and residential areas using the mosaic tool retaining maximum values (ESRI 2011). The resultant raster presented the cumulative potential impact of grazing, mining and infrastructure development by adding the weighted values. The highest values are recorded in areas where several threatening processes occur together. Areas located within the existing conservation estate and with steep slopes were combined with the distance grids by adding the negatively weighted rasters to the positively weighted combined disturbance raster. This modified the high values for distance to water and infrastructure if they were in a national park or on a steep slope. The values in the final condition layer were reversed so that "impact" values closer to zero were considered to have the lowest ecological value.

The merged dataset was then re-scaled to values between zero and one using the following formulae:

$Normalised\ (raster) = \frac{raster - Emin}{Emax - Emin}$, where *Emin* = minimum raster value, *Emax* = maximum raster value.

The combined and normalised raster illustrates the potential threatening processes from the best available data where values closer to zero represent the lowest possible ecosystem function whilst those closer to one representing areas with the highest possible ecosystem function.

Table 13. Relative weighting of disturbance impact and relationship to threatening process for habitat condition mapping.

VARIABLE	WEIGHT	RELATIONSHIP TO THREATENING PROCESS
Distance to water: 500m	15	
Distance to water: 1km	5	Weighting reflecting the level of damage with increasing distance from water, documented in Pringle and Landsberg (2004).
Distance to water: 4km	3	
Distance to water: 10km	2	
Distance to mining: 500m	30	Mining infrastructure has a clear impact on biodiversity but is likely to only influence the area directly at the site.
Distance to mining: 1km	10	
Distance to Primary road paved: 1km	7	Primary roads have the highest level of traffic, greatest diversity of users therefore the most likely vector for weeds and fire.
Distance to Primary road unpaved: 1km	6	

VARIABLE	WEIGHT	RELATIONSHIP TO THREATENING PROCESS
Distance to sec road paved: 1km	5	
Distance to sec road unpaved: 1km	4	Secondary roads have less traffic, but are still a major vector for weeds and fire. Requires access to water and borrow pits for maintenance.
Distance to other road paved: 1km	3	
Distance to other roads unpaved: 1km	2	
Distance to rail: 1km	5	Rail is a fixed path with a single user, so impact is less likely to spread post construction of rail corridor.
Distance to town <1000 people: 1km	10	Areas with smaller populations have less external infrastructure to support the town.
Distance to town <1000 people: 5km	5	
Distance to town >1000 people: 1km	20	Bigger towns have greater impact from transport and infrastructure.
Distance to town >1000 people: 5km	10	
Inside National Park	-2	Assuming a level of legislative protection that excludes extensive grazing, mining and other threatening process. Some demonstrated positive outcomes for terrestrial biodiversity in Australian rangelands (Kutt & Gordon 2012).
Slope greater than 30%, ridge tops and hill tops	-5	Demonstrated reduction in grazing probability on steep slopes and areas that are difficult or impossible to access for grazing animals (Prins & van Langevelde 2008).
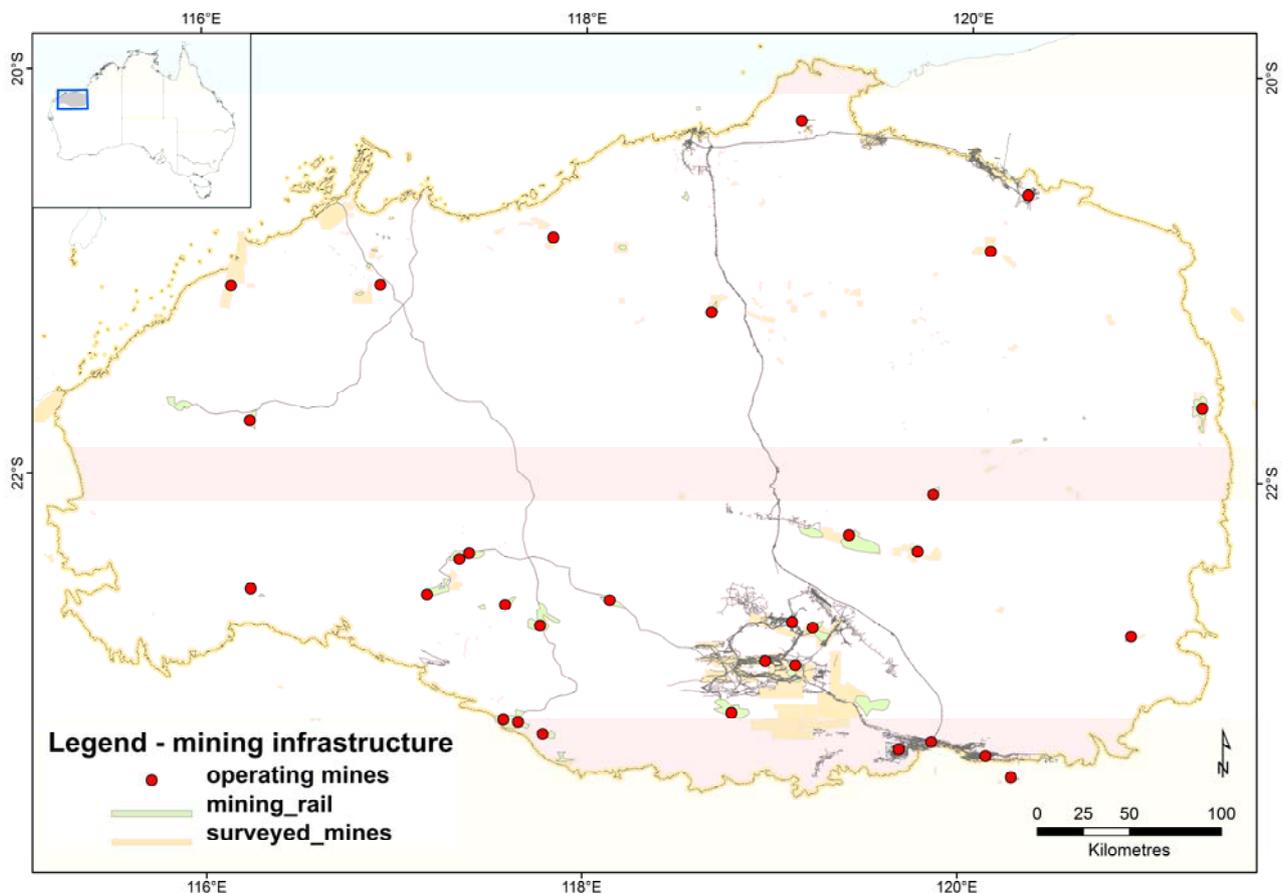
Figure 31. Combined mining infrastructure map for the Pilbara region. Major roads are also shown.

4.3 Results

Distance to water has been used as a proxy for grazing pressure and long term grazing pressure has been shown to transform floristic structure with long term consequences for the retention of overall biodiversity (Ludwig *et al.* 1999; Kutt *et al.* 2012). Figure 32 illustrates the historical grazing pressure from fixed water points and indicates areas with high density of watering points (highest grazing pressure – dark red) and those areas that have never been under significant induced grazing pressure (blue).
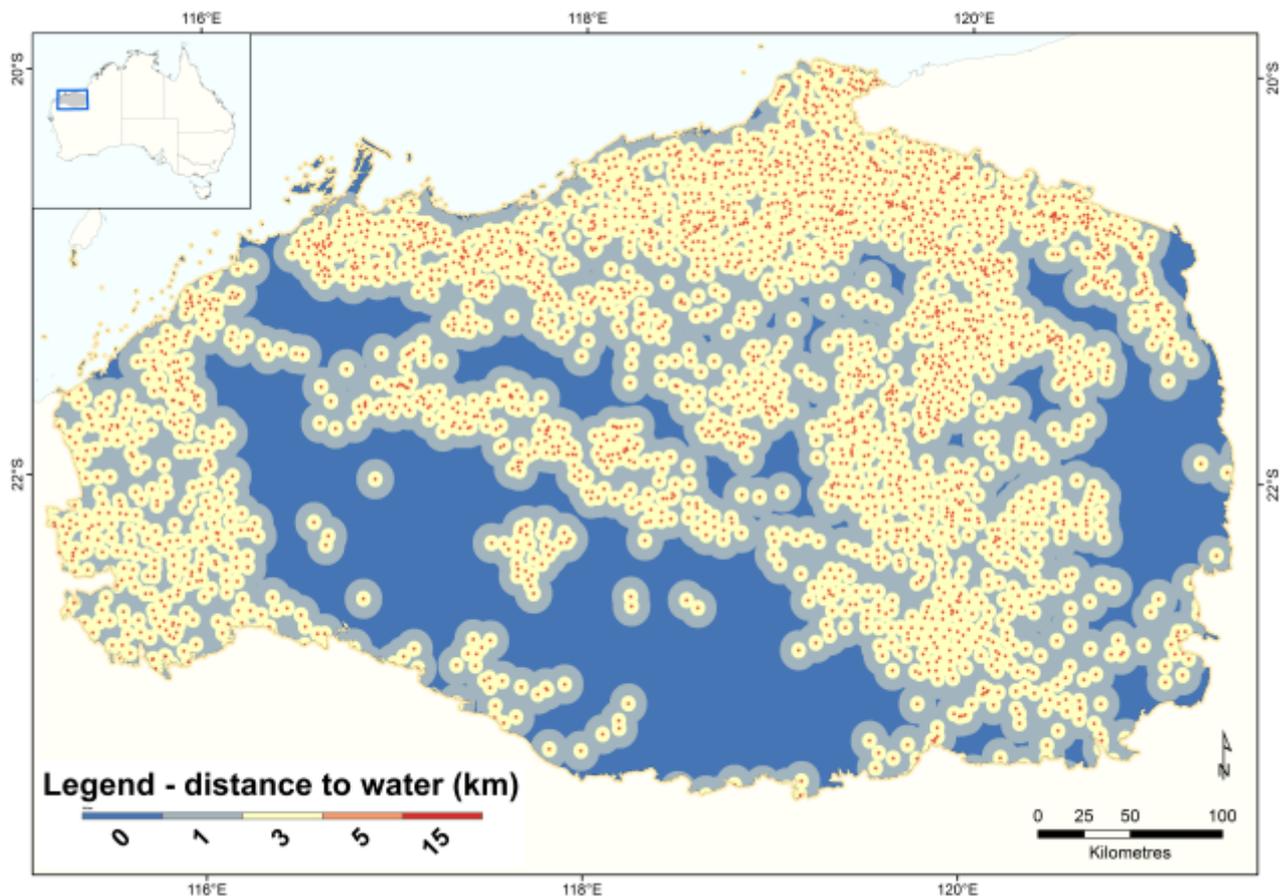
Figure 32. Distance to water points (dams, bores and windmills 1949/50) (decimal degrees, 0.0025 cell resolution).

A combination of best-available proxy variables for disturbance was used to derive an interim potential biodiversity habitat condition map (Figure 33). This map includes the effect of historical grazing pressure (1949/50 water points), contemporary pressure (mining, residential and transport infrastructure – listed in Table 13) and ameliorating conditions (existing conservation reserves and steep slopes) (Figure 33). The use of historical grazing pressure indicators illustrates the long term impact, and while more recent mapping of watering points is desirable, accurate data for contemporary watering points is not presently available.

Analysis of cattle movement data has demonstrated that the most predictive models for stock distribution, approximating potential grazing pressure, include terrain and distance to water (Prins & van Langevelde 2008). There are clear negative impacts on stock distribution and utilisation on areas with slope greater than 10% (Bailey 2005). Beyond 10% slope utilisation rates are dramatically reduced sharply declining above 30% slope and areas are effectively un-grazed beyond 60% slope (Holechek 1988). Landscape position is also likely to impact on the historical impact of grazing with areas such as plateaux surrounded by steep slopes providing an impenetrable barrier to stock.

Generally speaking, the Pilbara region is sparsely populated and the most significant impact on biodiversity habitat condition has been the introduction of grazing stock, mainly on the alluvial plains which provide the most accessible and productive pastures. The vast sand plains on the western edge of the Pilbara region are the least developed region (Figure 34). Historically, the rocky regions were also sparsely utilised by the pastoral industry, but mining development is now widespread with generally localised impacts resulting in similar levels of land use pressure as the alluvial plains (Table 14, Figure 33).
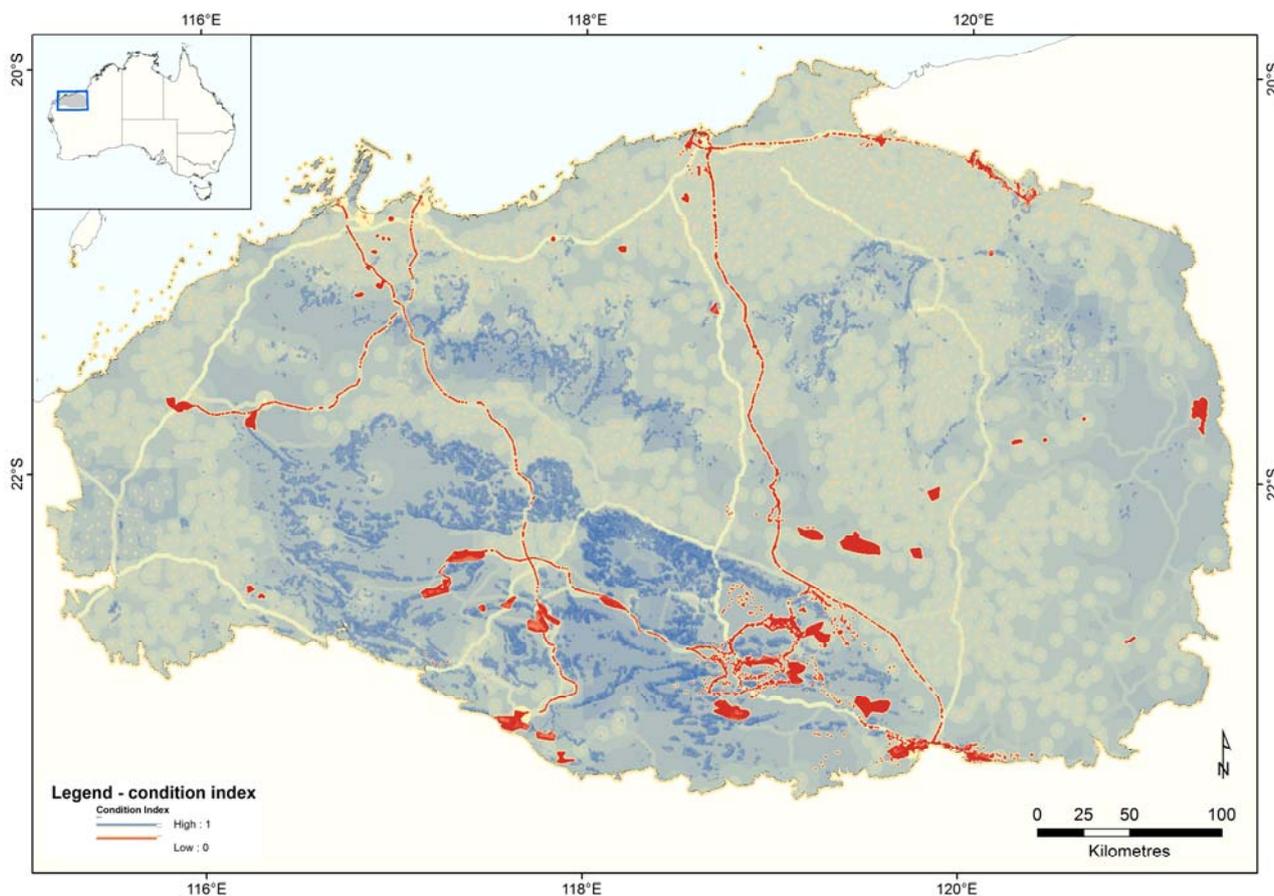
Figure 33. Interim condition grid illustrating areas of potentially low ecosystem function (red) contrasting with areas of higher ecosystem function (dark blue) as defined by distance to potential threatening processes.

Table 14. Mean, standard deviation and median ecological values for the dominant regolith types in the Pilbara region (Figure 34) in order of lowest impact (closer to 1) to highest impact (closer to 0).

REGOLITH TYPE	MEAN	STD	MEDIAN
Sandplain	0.76	0.06	0.77
Calcrete	0.74	0.07	0.77
Exposed	0.74	0.10	0.75
Residual	0.73	0.11	0.77
Colluvium	0.71	0.11	0.73
Alluvium	0.70	0.11	0.71
Anthropogenic areas	0.52	0.29	0.71
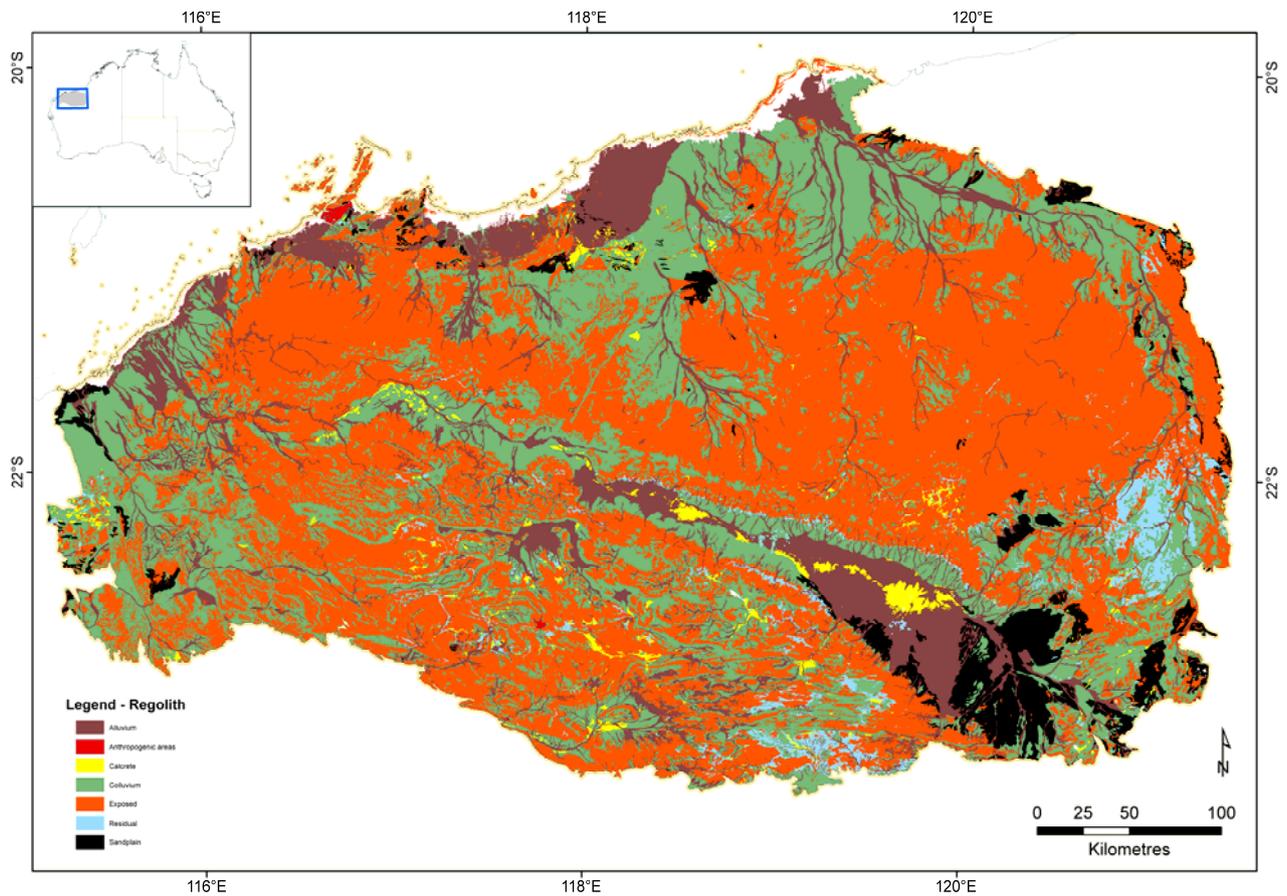
Figure 34. Major regolith classes in the Pilbara bioregion (Marnham & Morris 2003).

4.4 Discussion

We investigated potential data sources and techniques for remote mapping of habitat condition across the Pilbara. Two approaches were investigated and trialled; one based on recent advances in remote sensing (reported in Appendix E) and the other based on best-available proxy variables relating to various pressures (grazing, mining, residential and transport infrastructure) and ameliorating factors (protected areas and steep slopes). Contrasting with a traditional remote sensing definition of condition, which considers change between images over short periods of time, our definition of habitat condition aims to take into account the intrinsic capacity for biodiversity to persist dynamically over longer time scales in the presence of natural disturbances. This definition is designed to work flexibly with a biodiversity assessment and modelling framework for estimating the overall regional retention of native biodiversity (Zerger *et al.* 2013). The quantification and separation of natural and anthropogenic drivers of disturbance are therefore critical to this assessment.

As outlined in a review of the supporting literature and data sources presented in this section, the quantification of habitat condition at any scale is a very demanding task. While demonstrating the potential of remote sensing for assessing habitat condition (Appendix E), we also recognise the greater effort required to achieve this task with credibility. Therefore, in order to fully demonstrate how condition information is used in biodiversity assessment, we developed a model of condition based on proxy variables and inference from the literature (Figure 33). This interim estimation of habitat condition incorporates a number of data layers where infrastructure has been mapped. More explicit mapping and ranking of disturbance impacts, such as presently under development by BHPBIO for the strategic regional assessment, could also be incorporated into this estimation of habitat condition. All sources of disturbance information are also potentially useful verification and calibration data for a remote sensing application.

In review of the process and other work currently under development by CSIRO (Donohue *et al.* 2013), we suggest the following areas for further development and refinement to achieve a condition assessment model that is fully integrated with remote sensing, and as a baseline for cumulative assessment of impacts:

1. Formalise the theoretical framework for condition assessment. A theoretical framework for measuring condition in the context of biodiversity would involve; a) identifying threats to biodiversity; b) identifying spatially and temporally where those threats are located in the region; c) identifying the impact of the threats on aspects of biodiversity across the region (biodiversity response to disturbance); and d) conduct field validation surveys by identifying a gradient of disturbance and measure the differences in biodiversity across the gradient.

2. Image capture and processing. There is a considerable volume of data from a variety of satellite missions, time periods, scales and resolutions. A full assessment of the available data, derivatives and processing gaps is needed, including staying abreast of image capture and processing of Landsat 8 for near real time dynamic modelling.

3. Building a condition assessment and modelling pipeline. Critical to the assessment of habitat condition is a way to classify environments with similar characteristics with respect to their remote sensing signature, calibrated by their condition based on *in situ* observations – essentially establishing a field protocol for biodiversity condition assessment coupled with remote sensing. An initial framework to achieve this goal was recently developed by Donohue *et al.* (2013) who demonstrated the potential through a continental case study. This is akin to deriving a land system classification, except on a continuous scale. Reference sites to capture the natural variability across the region could be positioned evenly in environmental space where the condition is relatively pristine. These reference sites could then be used to measure fluctuations in biodiversity response characteristics (e.g., vegetation cover; bare ground) to define the influence of temporal variation in rainfall and natural disturbance regimes, for example. Other areas of varying condition due to anthropogenic drivers and of similar environmental profile can then be compared with the reference condition sites (i.e., is ground cover higher, lower or the same as the reference sites). Adjusted cover indices can be derived to illustrate areas where cover has been persistently above, below or similar to that of the reference sites; thus distinguishing areas where impacts are accruing for further investigation.

4. Rapid assessment. There needs to be a comprehensive rapid assessment of values to ground truth any condition model output and provide management context. This could include aerial surveys of domestic stock and other large introduced herbivores such as goats and camels. The surveys could be supplemented by distance sampling techniques to represent the effect of stock and feral animal densities across the region, and their varying impact on different land types. The region could be stratified into areas that have very high, very low and normal cover and biodiversity surveys conducted representatively to validate the remotely sensed indices for predicting biodiversity habitat condition.

5. Ongoing assessment of cumulative impacts. For effective monitoring, high resolution imagery (e.g., Spot5 2. 5m) could be used to derive up to date infrastructure maps and to provide a more accurate representation of contemporary water points that concentrate grazing pressure. Such imagery interpretation will provide a much finer grained perspective on the sources of disturbance without requiring extensive physical ground truth sampling. For example, Spot 5 imagery could also provide a high resolution map of bare ground for the region; potentially providing a fine resolution picture of the extent of anthropogenic disturbance at the extreme end of the condition spectrum.

6. Fire history mapping. Altered fire regimes (departing from a natural regime) in an arid to semiarid environment can have serious consequences for resident biodiversity. Regular mapping of fire scars using MODIS and Landsat imagery will allow areas to be identified where a rapid decrease in ground cover has occurred, and attributed to natural or anthropogenic sources, that may otherwise skew condition model outputs.

We therefore recommend an initial trial of an integrated biodiversity habitat condition assessment system that utilises the framework developed by Donohue *et al.* (2013) comprising two parts – a) an initial "one-off" assessment using remote sensing data with sufficient rigour to inform the regional assessment process through the biodiversity significance model; b) an enhanced model, potentially with temporal dynamics, improved underpinnings, integrating different remote sensing data at highest possible resolution, with field verification and strategic monitoring to iteratively inform cumulative impact assessment.

5 Biodiversity significance analysis

5.1 Introduction

The modelled patterns of richness and compositional turnover (from Section 3) are here brought together with the interim condition layer (from Section 4) to estimate, and map, the "biodiversity significance" of each and every location (approximately 250m grid cell) across the bioregion. For the purposes of this report, biodiversity significance is assessed as the potential for a given location to harbour a concentration of species narrowly distributed beyond that location, due to natural patterns of endemism and/or anthropogenic habitat degradation. The general approach that we have used to estimate this potential is one that has been applied widely over the past 10 years, and has been purposely designed to work effectively with spatial layers of richness, compositional turnover, and habitat condition, all expressed on continuous scales of measurement. For further background to, and explanation of, this general approach see Ferrier *et al.* (2004), Allnutt *et al.* (2008), Ferrier *et al.* (2009), and Ferrier and Drielsma (2010).

5.2 Methods

The measures of biodiversity significance presented here were derived separately for each of the five biological groups (mammals, birds, reptiles, invertebrates, vascular plants) as a function of three component variables:

s_{ij} = the compositional similarity (proportion of species shared) between grid-cells i and j expected if the habitat at both of these cells were in good condition. This similarity is predicted using the GDMs fitted in Section 3.1.

r_i = the richness (number of species) at grid-cell i expected if this cell were in good condition. This is predicted using the GAMs fitted in Section 3.2.

c_i = the estimated condition of habitat (0 to 1) at grid-cell i based on the interim habitat condition layer developed in Section 4.

Given various concerns regarding the modelling of richness (discussed in Section 3.2, page 38), and the mapping of habitat condition (discussed in Section 4.4), six different variants of the biodiversity significance measure were derived for each group. The first three of these variants excluded any consideration of modelled richness, and therefore assessed significance in terms of compositional turnover alone.

The first of these three measures of biodiversity significance excluding richness, also excludes any consideration of habitat condition – i.e., the measure treats all locations (grid-cells) across the region as if they were in pristine condition. The significance of a given grid-cell is then assessed by using the species-area relationship[1] (with an exponent of 0.25) to estimate the proportion of species diversity occurring, collectively, across all grid-cells in a similar environment (i.e., cells predicted to be compositionally similar as a result of scaling the environmental space by the relevant GDM) to the cell of interest (*i*), that would be lost if the habitat of that cell were removed from the system (see Ferrier *et al.* (2004) and Allnutt *et al.* (2008)):

---

[1] For a recent discussion of the species-area relationship applied in this context, see Rybicki and Hanski (2013) and Faith *et al.* (2008).

Equation 1. Biodiversity significance (natural uniqueness)

$$BS_i = \left\{ 1 - \left[\frac{\left(\sum_{j=1}^{n} s_{ij} \right) - 1}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\}$$

This measure of biodiversity significance (*BS,* Equation 1) therefore assesses the <u>natural</u> uniqueness, or level of endemism, of the biodiversity that occurs, or would once have occurred, at a given location. Removing a grid-cell (represented by "-1" in the above equation) from a naturally rare environment (i.e., one likely to contain species with narrow distributions) will result in greater loss of regional biodiversity than removing a cell from a naturally extensive environment.

The second measure of biodiversity significance extends the above approach to consider not just the natural rarity of an environment (and its associated biodiversity) but also the extent to which the condition of habitat throughout this environment has been degraded through human disturbance:

Equation 2. Biodiversity significance (incorporating regional variation in habitat condition)

$$BS_i = \left\{ \left[\frac{\sum_{j=1}^{n} s_{ij} c_j}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} - \left[\frac{\left(\sum_{j=1}^{n} s_{ij} c_j \right) - 1}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\}$$

This measure (Equation 2) again uses the species-area relationship to estimate the proportion of species diversity occurring, collectively, across all grid-cells in a similar environment to the cell of interest, that would be lost if a full grid-cell's worth of habitat were removed from the system. However, this cell is now removed from a system in which habitat, and therefore biodiversity, has already been lost from the environment of interest. The proportion of species diversity remaining before the removal of the cell is estimated in the above equation by:

$$\left[\frac{\sum_{j=1}^{n} s_{ij} c_j}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25}$$

It is important to note that this second measure of biodiversity significance (Equation 2) considers habitat condition only in a particular way. It uses the interim condition layer (developed in Section 4) to estimate the effective proportion of habitat remaining in cells with a similar environment to the cell of interest (which we refer to as "regional condition" in the remainder of the report), but does not consider the condition of the cell itself (referred to hereon as "local condition") when removing this cell from the system – i.e., the cell is here treated as if it were in pristine condition.

Replacing "-1" in the above calculations with "-c_i" yields the third measure of biodiversity significance:

Equation 3. Biodiversity significance (incorporating the effects of regional and local variation in habitat condition)

$$BS_i = \left\{ \left[\frac{\sum_{j=1}^{n} s_{ij} c_j}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} - \left[\frac{\left(\sum_{j=1}^{n} s_{ij} c_j \right) - c_i}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\}$$

Like the second measure, this estimates the proportion of species diversity that would be lost if a given cell were removed from the system, but in doing so derives the habitat condition of this cell from the interim condition layer rather than treating it as if it were in pristine condition. This measure (Equation 3) therefore considers both "regional and local condition".

The remaining three variants of biodiversity significance are derived by scaling each of the three measures described above according to the predicted richness of the environment concerned, by multiplying by:

$$\frac{\ln(r_i)}{\ln(r_{max})}$$

where r_{max} is the maximum richness, anywhere in the Pilbara, predicted by the GAM for this biological group. These last three measures are therefore:

Equation 4. Biodiversity significance (natural uniqueness scaled by species richness)

$$BS_i = \left\{ 1 - \left[\frac{\left(\sum_{j=1}^{n} s_{ij}\right) - 1}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\} \frac{\ln(r_i)}{\ln(r_{max})}$$

Equation 5. Biodiversity significance (regional variation in habitat condition scaled by species richness)

$$BS_i = \left\{ \left[\frac{\sum_{j=1}^{n} s_{ij} c_j}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} - \left[\frac{\left(\sum_{j=1}^{n} s_{ij} c_j\right) - 1}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\} \frac{\ln(r_i)}{\ln(r_{max})}$$

Equation 6. Biodiversity significance (regional and local variation in habitat condition scaled by species richness)

$$BS_i = \left\{ \left[\frac{\sum_{j=1}^{n} s_{ij} c_j}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} - \left[\frac{\left(\sum_{j=1}^{n} s_{ij} c_j\right) - c_i}{\sum_{j=1}^{n} s_{ij}} \right]^{0.25} \right\} \frac{\ln(r_i)}{\ln(r_{max})}$$

5.2.1 PRESENTATION OF RESULTS WITH UNCERTAINTY INCLUDED

Details of the algorithms defining uncertainty associated with the community-level modelling from the applications of GDM (species compositional turnover) and GAM (species richness) are described in Chapter 3. In addition, here we describe the parameters used to map these "uncertainty clouds".

GDM model uncertainty

For each biological group, the GDM uncertainty was calculated as a function of the proportion of the density of survey effort within GDM scaled environmental space (as defined by the equation in Section 3.1, page 20). An arbitrary threshold of 0.25% was used to define certainty (i.e., a value of 1), such that all cells whose scaled environment had been surveyed to an extent of >0.25% were considered adequately sampled. Certainty was linearly scaled to uncertain (i.e., a value of 0) at 0% sampling (see Appendix F for each biological group – Figures F16, F26, F36, F46, F56). The uncertainty cloud for the average Biodiversity Significance across all biological groups was calculated as double the arithmetic mean of the five contributing layers to ensure an appropriate level of opacity. The arbitrary 0.25% threshold here refers to the percentage of similar GDM-scaled environmental space across the whole of the Pilbara which has been sampled, rather than the proportion of the cell itself. Thus we are dealing with a continuous measure of similarity of environments. We assume that any cell that has been surveyed has been adequately surveyed and do not attempt to estimate within cell survey error.

GAM model uncertainty

Uncertainty in the GAM richness model was mapped as a function of two factors. Firstly, all modelled values which fell outside the range, [1-(maximum observed richness)+10%] (i.e., those which were capped in Appendix F Figures F15, F25, F35, F45 and F55) were given a certainty value of 0 (i.e., uncertain). Secondly, the proportional standard error (SE) was calculated as [SE/richness] for each cell (see Appendix F for each biological group – Figures F17, F27, F37, F47 and F57). Values falling within an arbitrary cut-off point of 10% proportional standard error were assigned total certainty/transparency, with the uncertainty/opacity increasing linearly to full uncertainty at >25%. In practice, the two estimates gave similar results in most places. As for GDM, the uncertainty cloud for the average Biodiversity Significance across all biological groups was calculated as double the arithmetic mean of the five contributing layers to ensure an appropriate level of opacity. This is not to be confused with the 0.25% GDM scaled survey certainty described above. Here (GAMS) we are dealing with proportional standard error. Thus a value of 0 means there is no standard error, and increasing values indicate increasing standard error and therefore

increased uncertainty. We arbitrarily set the areas of high certainty to be transparent (<10%) and areas of high uncertainty (>25%) to be opaque.

Mapping uncertainty

When presenting the results, the relevant uncertainty layers (GDM and/or GAM) with variable transparency always overlayed the Biodiversity Significance. The layers are transparent where certainty is high, becoming increasingly opaque with increased uncertainty; thus, masking any results with lower than credible certainty. To achieve this effect, uncertainty layers (0-1) were converted to a 0 (certain) to 255 (uncertain) scale to represent the RGB/Alpha values respectively in a composite bands raster. The single layer of continuous value 255 for white (or 128 for grey) was used to define the overlay colour (Alpha values). If only one uncertainty layer is applicable, then the white tones are used (i.e., Equations 1 to 3). In cases where both the GAM and GDM uncertainty are applicable then the white tones signifies GDM uncertainty and grey signifies GAM uncertainty (i.e., Equations 4-6). In the Composite Bands function of ArcMap the two rasters were added as separate bands. The single continuous value layer (e.g., with values 1-255 for white) was allocated to the Red, Green and Blue bands, whilst the uncertainty layer (with values 0-1) was set as the active Alpha band. Values were stretched linearly, using minimum-maximum values, and the colour ramp inverted to give white. For the grey ramp, no inversion was required since the colour is always 50% grey. The result was saved as a single multi-band raster layer. Uncertainty overlays for GDM and GAM (richness) were saved and applied as separate files, which can be superimposed either individually or together (using the grey variation of the GAM uncertainty overlay) over any map.

5.3 Mapped outputs

Relative biodiversity significance (*BS*) values, derived using all six variants of the biodiversity significance measure (Equation 1 to Equation 6), are mapped across the entire Pilbara Bioregion for each of the five biological groups in Appendix F . These values are averaged across the five groups to provide an indication of overall significance, according to each of the six measures. The first three measures, excluding richness are presented in Figure 35 to Figure 37. The remaining three variants are scaled by relative richness (Figure 38), resulting in the alternative measures of biodiversity significance (Figure 39 to Figure 41).

The maps depict both the estimated level of biodiversity significance, from lower significance in green through to higher significance in red, as well as the relative level of certainty associated with these estimates, ranging from paler white (GDM uncertainty) and/or greyer (GAM uncertainty) colours in areas of lesser certainty (appearing washed out) through to brighter colours in areas of higher certainty (transparent).

It should be noted that each of the three approaches adopted for addressing condition in these Biodiversity Significance analyses is intended to convey a different message, and therefore requires a different interpretation:

- Maps labelled as "excluding condition" (Figure 35 and Figure 39) convey the relative significance of locations purely in terms of the natural level of uniqueness, or endemism, they are expected to have exhibited prior to anthropogenic disturbance.
- Maps labelled as "including regional condition" (Figure 36 and Figure 40) additionally factor in the effects that habitat degradation throughout the region has had on the amount of habitat, and associated biodiversity, remaining in different types of environments. The interim condition layer (from Section 4.3) is here used purely to provide regional context, in terms of the overall level of habitat degradation within different environments, but not to assume anything about the condition of any particular location of interest (i.e., a cell in a given environment will exhibit a constant level of significance in this analysis regardless of its local condition).
- Maps labelled as "including regional and local condition" (Figure 37 and Figure 41) factor local condition (as mapped by the interim condition layer) directly into the analysis – i.e., within a given environment, cells in poorer condition will be mapped as having lower significance than cells in better condition.

The colour ramps used to map Biodiversity Significance are comparable across the two series (i.e., maps excluding richness based on Equations 1, 2 and 3; and maps including richness based on Equations 4, 5 and 6). This allows comparison of condition estimates (with or without condition), but the mapped colour ramps are not comparable between series (i.e., with and without richness included). For additional context, an overlay shows the location of conservation reserves (NRS 2010).

The overall or regional context in which condition is taken into account indicates how much of that 'environment' (i.e., cells predicted to be compositionally similar as a result of scaling the environmental space by the relevant GDM) has been lost or degraded as a result of human land use activities. The lower the condition value, the higher the grid cell of interest relative to the level of degradation of other places that are predicted to be compositionally similar. The local context in which condition is taken into account considers just the condition of that grid cell. The choice of measure depends on the purpose for which biodiversity significance is being assessed; for example, the contrast between development and restoration.

The maps presented in Appendix F include the GAM richness model and GDM and GAM uncertainty surfaces for reference. It is important to note that the GDM-based estimate of survey density and the GAM statistical estimate of richness standard errors are not comparable measures either conceptually or quantitatively. The two models use different survey points and different environmental variables and the uncertainty estimates were derived differently. However, both work in modelled environmental space. The GDM uncertainty is a measure of the sampling of GDM space similar to each cell, and is a simple estimate of sampling density, and is not related to statistical uncertainty. Areas of high sampling coverage generally correspond with higher confidence in GDM-modelled environmental space. The GAM standard error is a statistically-based estimate derived from the confidence in richness model fit.

The choice of colour ramps and classifications aim to maximise the information contained in the maps whilst maintaining comparability within map series. Thus the thresholds were set at a level which allowed the spatial detail of the maps to be seen across all similar maps and the relative properties of maps to be compared.
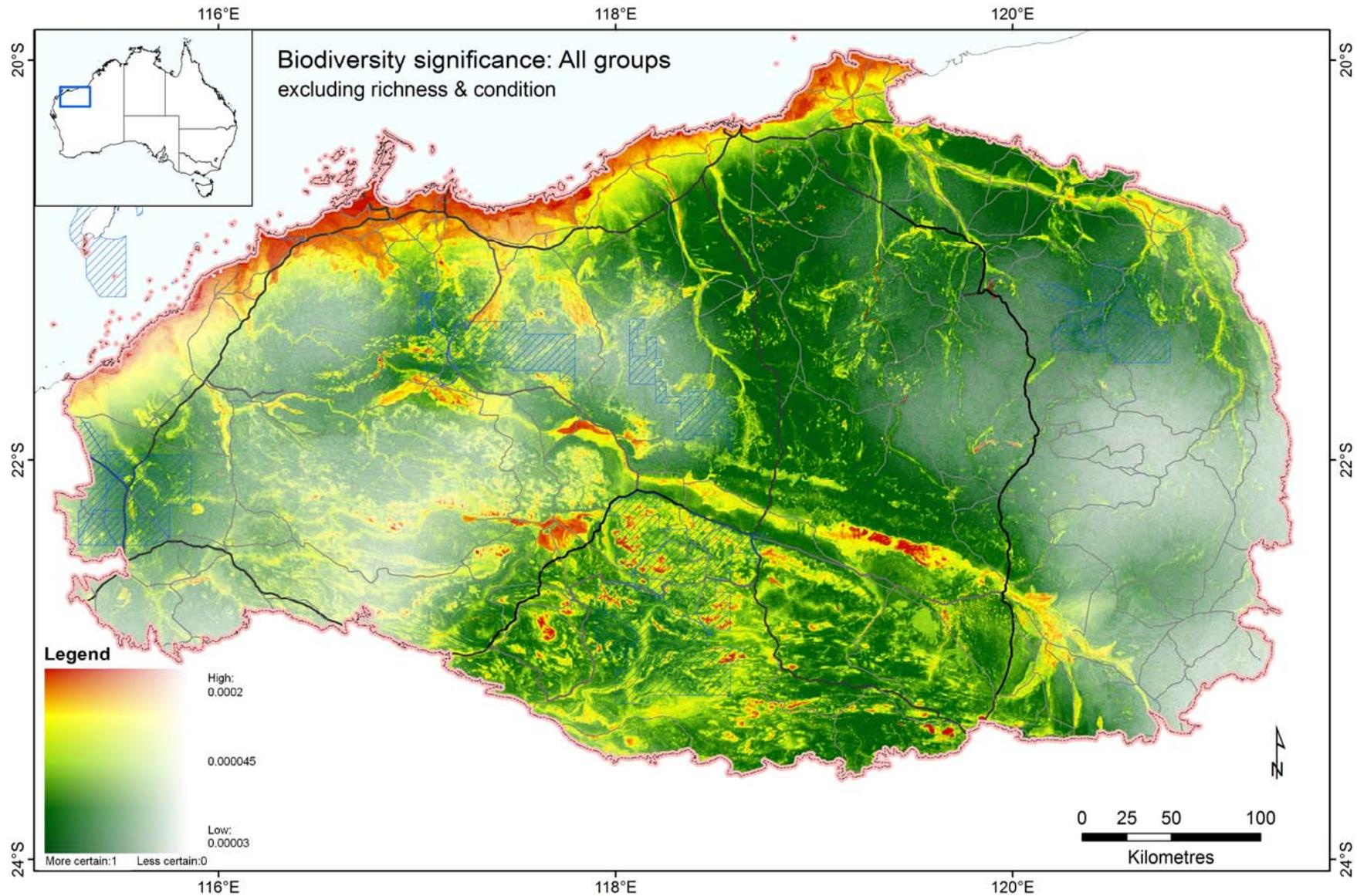
Figure 35. Biodiversity significance (from Equation 1), excluding richness and condition for all groups, based on community-level modelling across the Pilbara. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
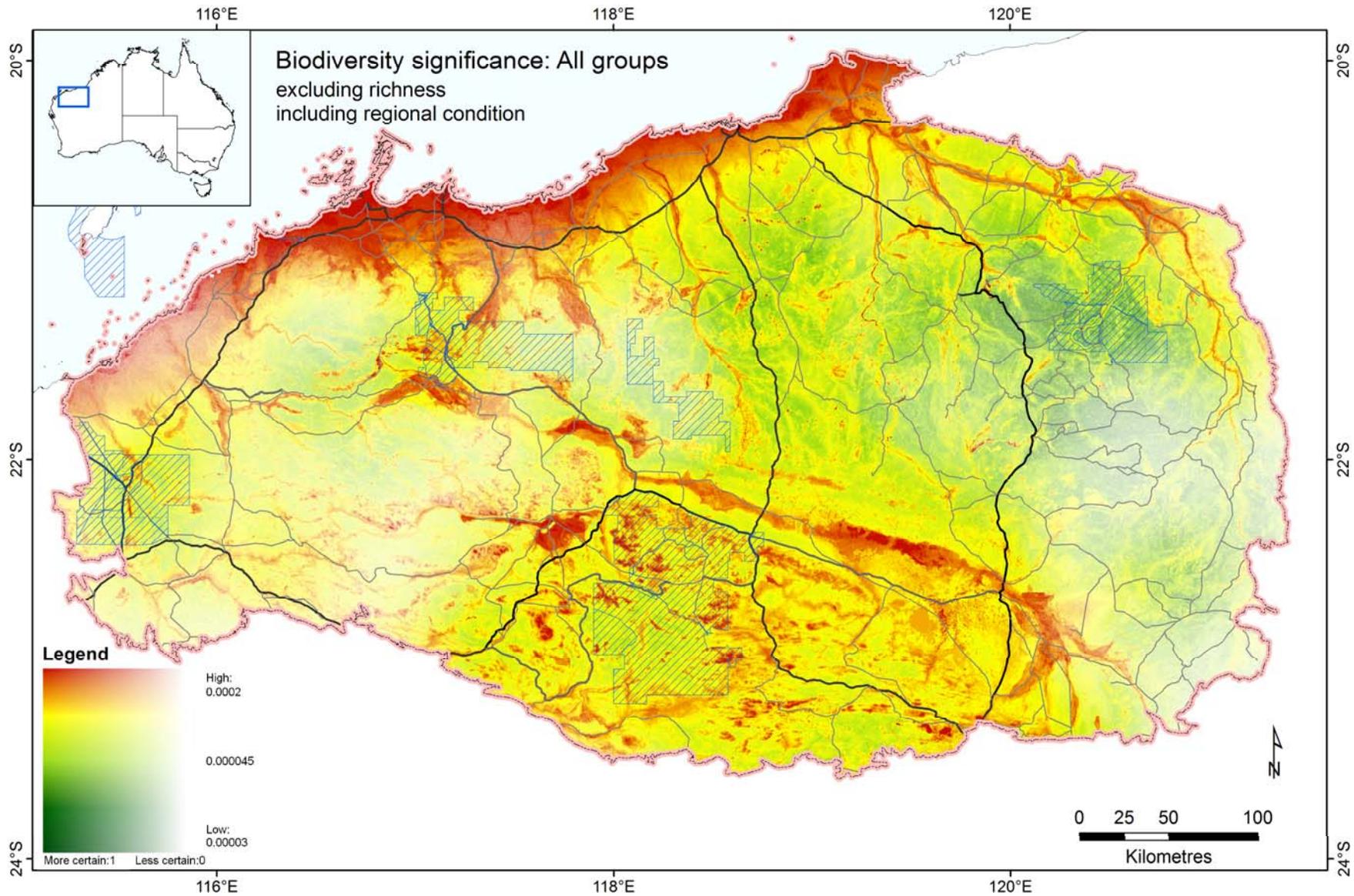
Figure 36. Biodiversity significance (Equation 2) excluding richness and including regional condition for all groups, based on community-level modelling across the Pilbara. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
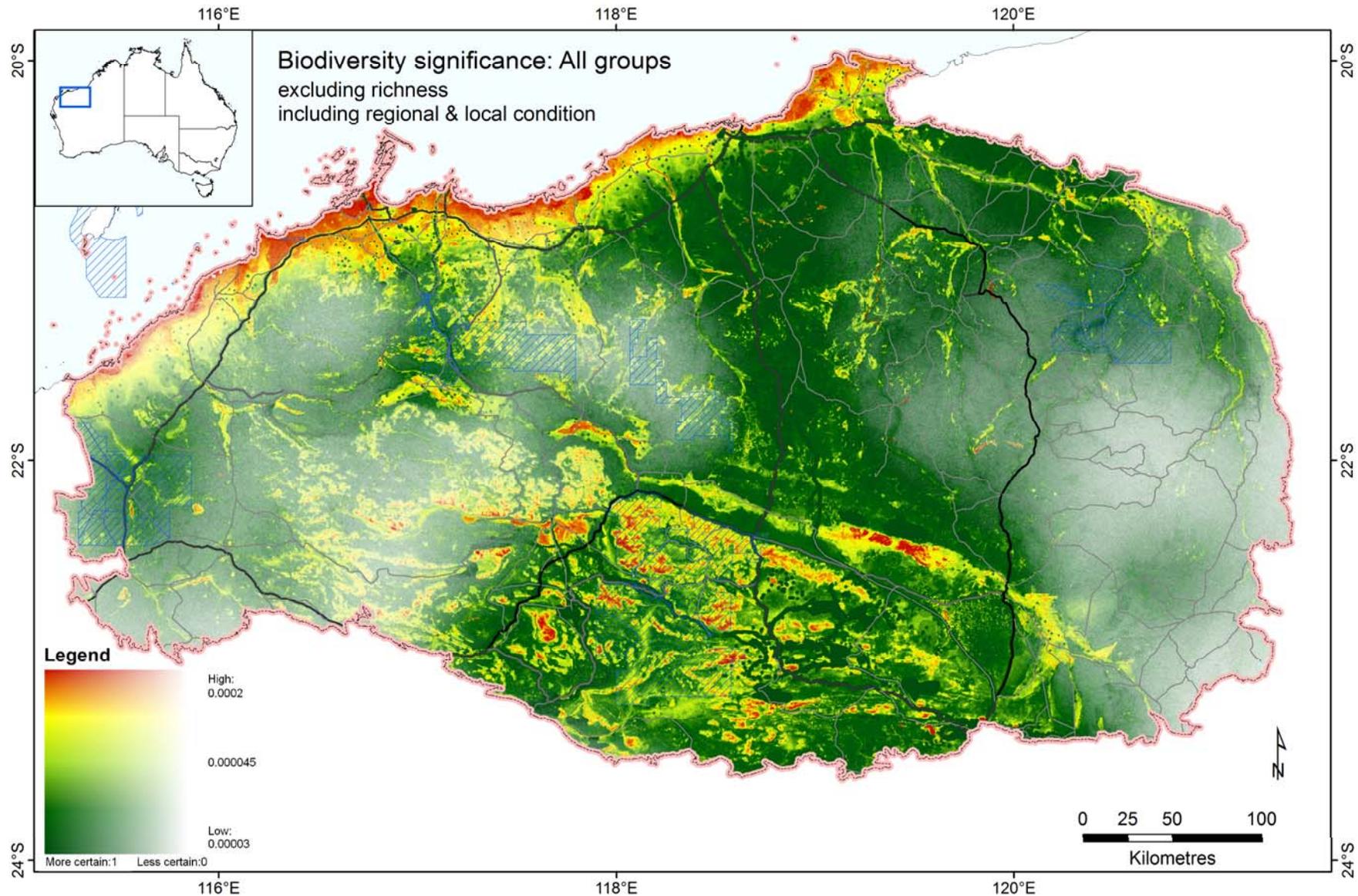
Figure 37. Biodiversity significance (Equation 3) excluding richness and including regional and local condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
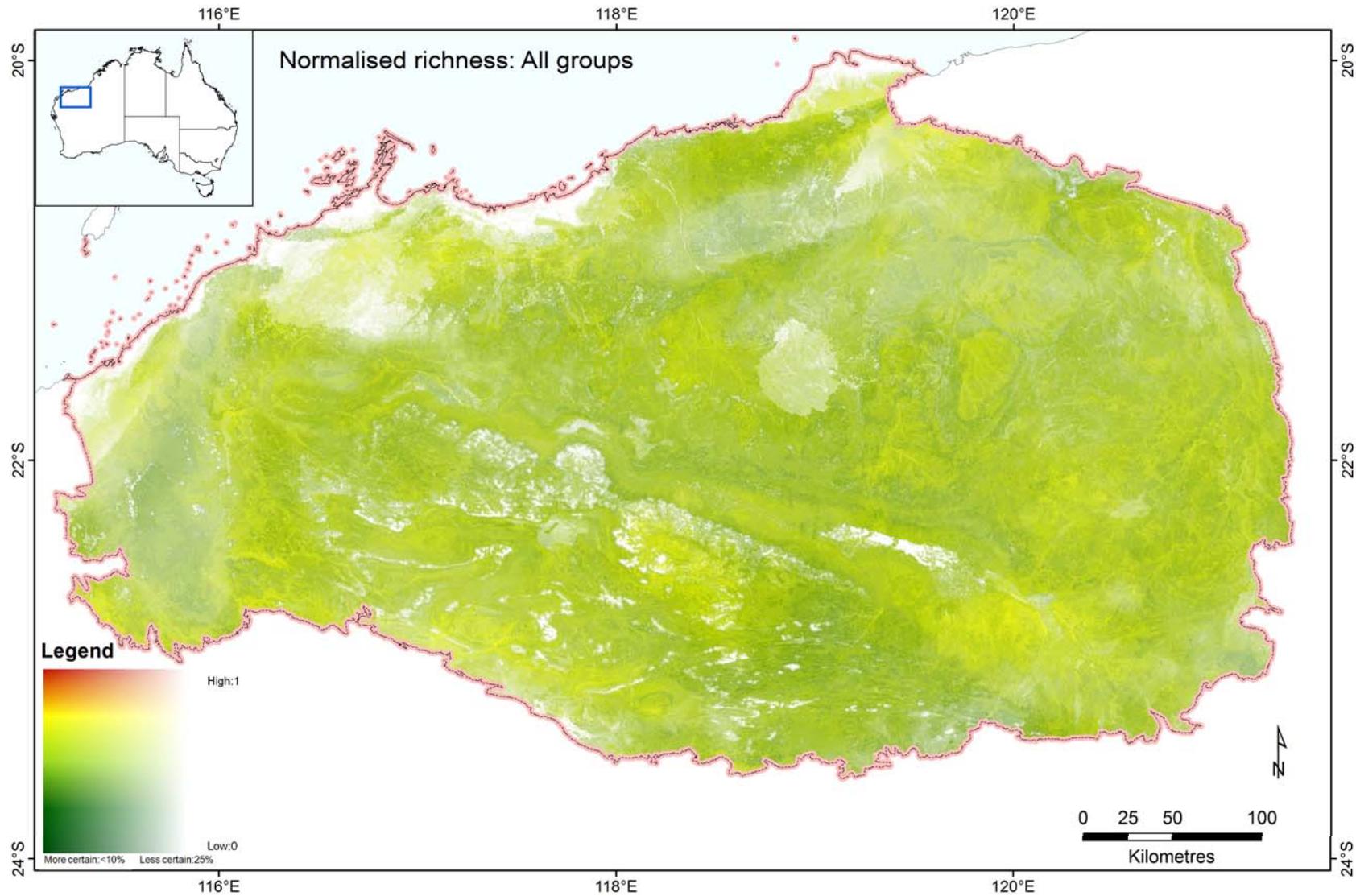
Figure 38. Normalised species richness (log fraction of maximum richness) averaged across the five biological groups, based on community-level modelling (Section 3.2). This is the right hand side, $\ln(r_i)/\ln(r_{max})$ of the biodiversity significance equation (i.e., used in Equations 4, 5 and 6). Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas (composite GAM uncertainty derived from the standard error of the predicted value in each case).
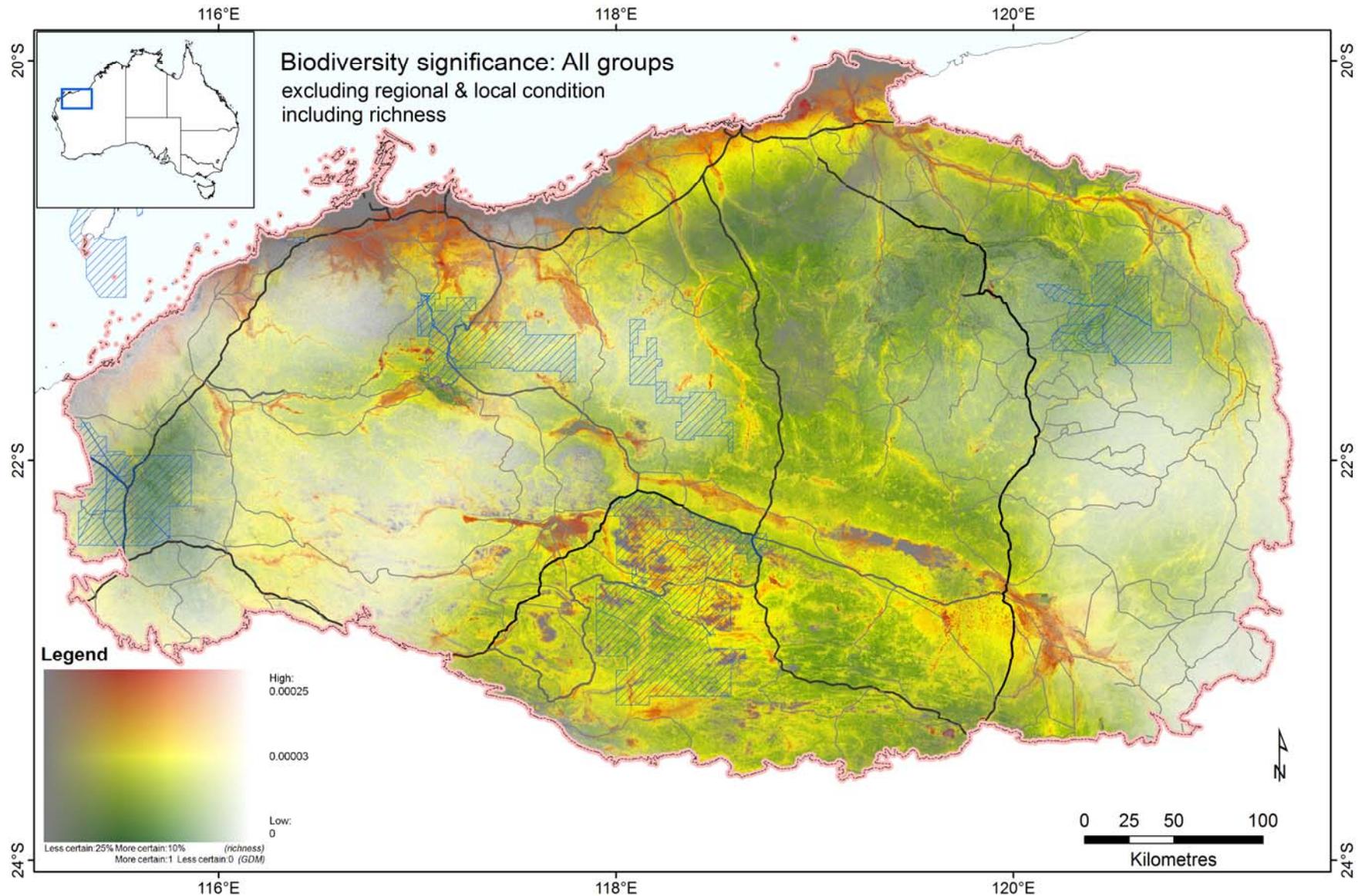
Figure 39. Biodiversity significance (Equation 4) including richness excluding condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.

Figure 40. Biodiversity significance (Equation 5) including richness and regional condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
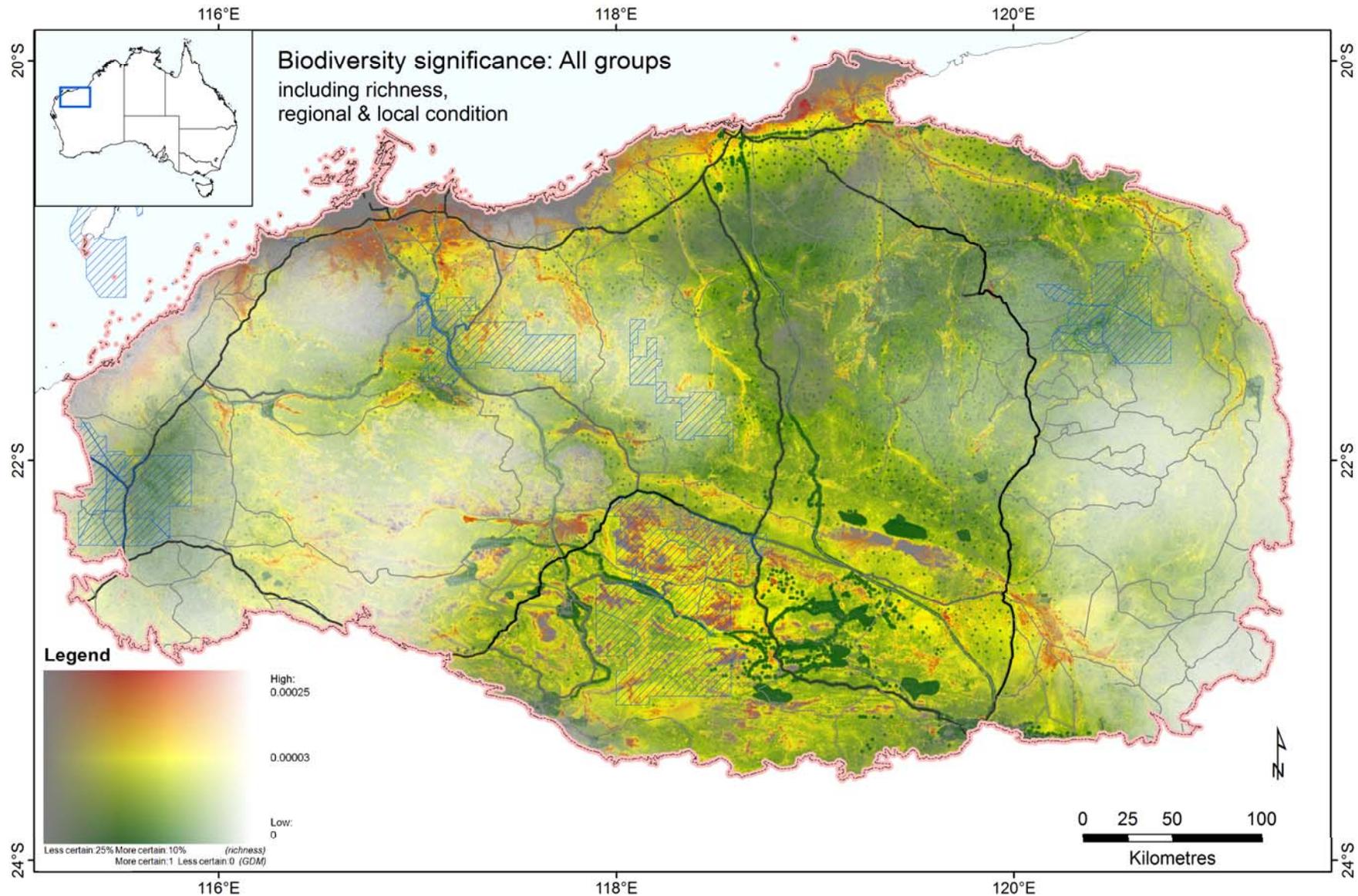
Figure 41. Biodiversity significance (Equation 6) including richness, regional and local condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.

5.4 Results and discussion

The Biodiversity Significance assessments represent the culmination of activities outlined in Figure 1. The base calculation of Biodiversity Significance uses the models of species compositional dissimilarity (also referred to as β-diversity in the literature) developed in Section 3.1. In the absence of estimates of habitat condition, Biodiversity Significance measures natural uniqueness (Equation 1, Figure 35). Compositional dissimilarity, however, is only a partial measure of emergent community-level properties of biodiversity. The scaling by species richness (e.g., Equation 4, and see Figure 38) enables an overall measure of expected biodiversity persistence to be generated (Ferrier *et al.* 2009), and so clarifies the level of natural uniqueness at one location compared with another (compare Figure 35 with Figure 39, and see Appendix F for individual biological groups). While both compositional dissimilarity and species richness (also referred to as α-diversity in the literature) are both relevant to the calculation of Biodiversity Significance, the β-diversity component is the primary measure because it defines how different or unique biodiversity is at one location compared with biodiversity elsewhere within the region (thereby underpinning the principle of biodiversity complementarity in systematic conservation planning, Margules & Pressey 2000). Locations rich in species will be more significant, but how significant a location actually is depends on whether the species are widespread elsewhere or only found at a few locations (i.e., uniqueness). Ideally we would use both α- and β-diversity components in the calculation of Biodiversity Significance, but given the problems we experienced in modelling species richness (discussed in Section 3.2), alternative calculations of Biodiversity Significance are presented for evaluation and potential use in regional assessments. The choice of Biodiversity Significance calculation depends on how well the models of species richness represent actual patterns in the landscape, which requires validation through comprehensive and systematic biological survey to fill gaps in existing survey coverage.

When regional estimates of habitat condition are taken into account, relative uniqueness is rescaled, thus highlighting remnant areas of regionally unique biodiversity relative to their original (natural or pristine) extent based on the models (e.g., see Figure 36 and Figure 40). The local influence of habitat condition can also be considered such that, within a similar type of environment, locations in poorer condition are mapped with lower significance than those in better condition (e.g., see Figure 37 and Figure 41; and compare with the condition map - Figure 33). While we have provided legends for the maps that allow comparison within each group (with or without richness), a different legend may be needed if the maps are used in isolation to ensure the maximum amount of information is conveyed to support decision making.

We also present Biodiversity Significance as an aggregate of the five biological groups by averaging the outputs across these groups; essentially treating them as equally important as each other. With three terrestrial vertebrate groups (birds, mammals and reptiles), one invertebrate group (comprising three groups) and one group comprising all vascular plants, this average is naturally weighted in importance toward the vertebrate fauna, and toward fauna overall. Alternative ways of weighting and aggregating the models for each biological group could be explored. Therefore choices can be made about using an aggregated estimate of Biodiversity Significance, or separate estimates for each biological group of relevance to the assessment process.

The separate presentation of the Biodiversity Significance calculations serves two objectives: 1) a demonstration of how the various input layers and models influence the estimation of Biodiversity Significance and, 2) given that further work is required to ensure the robustness of the richness and condition models, choices about the most appropriate Biodiversity Significance calculation can be made depending on the purpose and the relative influence each may have on critical assessment decisions.

While we recommend against placing too much reliance on outputs that have included richness; we feel more confident about the use of outputs that have included condition. While the outputs including also local condition (e.g., Figure 37) assume a level of site-specific reliability in the mapping of condition that requires validation, we consider the Biodiversity Significance analysis with regional condition (e.g., Figure 36) to be considerably more robust because these account for the overall (average) level of degradation across different environments within the region, and not condition at any particular site of interest. The

former is more applicable to a regional assessment process with regard to large-scale development decisions and the latter to incremental decisions such as restoration priorities or offsets and scenarios of cumulative impact.

Ideally, both forms of condition would be used in all land use decisions. Uses of regional and local condition are only distinguished here because decisions based on local condition estimates require validation. For example, assuming sufficient rigour in the estimation of local condition, both regional and local condition would be used in the calculation of biodiversity significance to inform decisions about where to place a new mine. This assessment would show the proportional loss of particular environments (i.e., scaled by compositional dissimilarity) as a consequence of the proposed development and their relative significance. An area of high significance might be identified for two reasons: 1) it is a place that remains in good condition while similar 'environments' of that type that have elsewhere been degraded throughout the region, or 2) that type of 'environment' is naturally rare throughout the region. The assessment implies that development of such places of high significance be avoided in order to maintain the regional level of biodiversity.

In combination, the "excluding condition" " (Figure 35 and Figure 39) and "including regional condition" (Figure 36 and Figure 40) maps (particularly those "excluding richness") convey an interesting, and seemingly coherent, story. In general, places that have the highest significance (relative to other parts of the region) in the "excluding condition" map (e.g., Figure 35) also have highest significance in the "including regional condition" map (e.g., Figure 36). But the significance of all places is, on average, higher in the latter map (everything becomes redder) (see also Appendix F for individual biological groups). This sensibly reflects the added effect of anthropogenic habitat degradation on the rarity / uniqueness of the region's environments, and the types of biodiversity these environments support.

We have provided (in digital form) the raw materials for generating an even wider variety of "biodiversity significance" measures than those presented here. The relative utility of these different measures depends on the particular question BHPBIO is interested in, and therefore the particular type of significance that is of relevance. Assistance in matching the relevant measure to the particular question requires more precise information about which decisions need to be informed by these maps or their underpinning models. For example, the "including regional condition" map (e.g., Figure 36) has considerable potential to inform identification of priority places both for protection (or exclusion of development) and for restoration or for offsets. These different purposes require different interpretations of the same underlying data, ideally factoring in on-ground knowledge of site-specific condition at any location of interest. For example, a location mapped as highly significant on the "including regional and local condition" map (e.g., Figure 41) should have a high priority for protection (exclusion of development) if it is in reasonably good condition, but a high priority for restoration if it is currently in poorer condition.

6 General discussion

6.1 Overview of results

This report presents the results of a consultancy undertaken by CSIRO to assess spatial patterns in the distribution of biodiversity, and associated levels of biodiversity significance, across the Pilbara Bioregion. The work employed state-of-the-art techniques for community-level modelling and biodiversity assessment, integrating best-available existing biological and environmental data for the region. The "overall biodiversity" perspective adopted in this study, through the use of community-level data and modelling techniques, is purposely intended to complement other recent work commissioned by BHPBIO focusing on modelling and assessment of individual species of particular conservation concern.

The overall approach of this study (Section 1.2) involved a series of activities (Figure 1) aimed at gathering the best-available information on the distribution of flora and fauna (Section 2), environmental covariates and habitat condition (Section 4), used in models of compositional turnover (Section 3.1) and species richness (Section 3.2) to estimate and map relative levels of biodiversity significance across the Pilbara bioregion (Section 5). Because of weaknesses in the modelling of species richness (outlined in Section 3.1, page 28) and in the preliminary assessment of habitat condition (outlined in Section4.4), we recommend that only the biodiversity significance models generated using the compositional turnover models, with or without regional condition (e.g., Figure 35, Figure 36; and see Appendix F for individual biological groups), be used to inform the current regional assessment process (outlined in Section 5). Although requiring validation as a basis for critical planning and decision making, we also present the individual models with these components (i.e., richness and local condition) integrated (Appendix F). With the improvements identified in the preceding sections of this report, fully integrated models to support dynamic local and regional decision making are feasible.

6.1.1 COMMUNITY-LEVEL AND HABITAT CONDITION MODELLING

In Section 2, we presented the methods used in compiling and assessing the fitness-for-use of data sourced from BHPBIO, the Atlas of Living Australia and arising from systematic biodiversity surveys (George *et al.* 2011). These observations are necessarily biased by purpose (e.g., BHPBIO interest areas), accessibility (proximity of habitat to roads), and survey effort (i.e., incidental or targeted observations versus systematic survey). The richness modelling required greater rigour in the choice of data, restricted to those sources derived through some form of systematic survey (Table 8), compared with the more relaxed assumptions in sourcing data for modelling compositional turnover (Table 6). In both cases, however, there is a notable bias toward surveys in the vicinity of the BHPBIO interest areas (Figure 2). This bias will influence the outcome of the community-level modelling and the estimates of Biodiversity Significance, in favour of providing greater clarity around the BHPBIO interest areas. The bias could potentially increase the significance of BHPBIO's interest areas relative to other parts of the Pilbara due to inflated estimates of local species richness (Figure 12) or greater uniqueness; due to (unknown) inadequate sampling of unique biodiversity in other areas of the Pilbara. However, this does not appear to be a major concern (compare upper and lower estimated richness maps for each of Figure 21 to Figure 25; and compositional turnover patterns in Figure 15 to Figure 19). A sparse, but sufficiently representative sample has resulted in a reasonable interpolation of species richness and turnover across the Pilbara. For example, where the greatest bias might be anticipated with vascular plants (see Figure 11) the classification of regional compositional turnover reasonably reflects the subregional classification of bioregions and therefore uniqueness of the component biodiversity (compare Figure 19 and Figure 3).

For this application we were unable to access the systematic surveys of vascular plants across the Pilbara (George *et al.* 2011) for which analysis and species identifications were incomplete (Stephen Van Leeuwen

personal communication). A partial sample of these data was accessed, however, through the Atlas of Living Australia where specimens have been lodged with the Western Australian Herbarium and identified to the species taxonomic rank. This supplementary data, along with riparian vegetation surveys (Pinder *et al.* 2009), helped balance the more intense sampling around the BHPBIO tenements.

We understand that other mining and pastoral companies throughout the Pilbara are also required to conduct surveys in accordance with their development requirements. Further compilation of these data or their public contribution through the Atlas of Living Australia or the Terrestrial Ecosystem Research Network AEKOS data portal will help fill gaps in knowledge about patterns of biodiversity distribution and therefore assist in focussing future survey effort.

Given the importance of understanding the reliability of these models to potentially inform planning and decision making, we also generated spatial estimates of uncertainty. For the richness models, we used a recent enhancement of generalised additive modelling procedures (GAM), which allows likelihood based smoothing estimation (Wood 2006) wherein predictions can be accompanied by standard errors based on the posterior distribution of the model coefficients (Marra & Wood 2012). For the compositional turnover models, in the absence of a formal estimation of model error, we used sampling density in the respective GDM-scaled environmental space as a relative indicator of the level of certainty associated with predictions. The relative uncertainty therefore associated with either the GDM or GAM models for each biological group are presented along with the mapped biodiversity patterns in Sections 3.1 and 3.2. These maps demonstrate the different spatial variation in uncertainty associated with each group and each model.

The estimates of uncertainty are presented in combination with the models of Biodiversity Significance to show those regions within which we have greater or lesser confidence in the modelled estimates. For this application we developed a mask overlay with increasing opaqueness in areas of lesser confidence and transparent in areas of greater confidence (e.g., Figure 35). The two contrasting error surfaces were combined by scaling relative to whiteness (GDM) or greyness (GAM) (e.g., Figure 39). Detail about the process by which this novel overlay was generated is provided in the metadata associated with the datasets (public access restricted to BHPBIO via the CSIRO Data Access Portal, https://data.csiro.au/dap/home).

In addition to presenting information about model uncertainty, these error surfaces more classically represent locations where additional sampling will significantly improve the representation of environments and therefore biodiversity, leading to more robust models (Ferrier 2002). An example of this type of application for expeditionary surveys, albeit applying a slightly different algorithm, was developed by Funk *et al.* (2005). Proposed new surveys for a particular biological group can be guided by continuous environmental space scaled by a GDM.

With respect to monitoring, for example to inform calibration of remote sensing indices to derive reliable measures of biodiversity habitat condition or quality, we outlined an integrated assessment approach in Section 4.4. Here we refer to a recent proof of concept developed by Donohue *et al.* (2013), which could be trialled across the Pilbara, building on the preliminary remote-sensing based assessment of condition presented in Appendix E and the interim estimation of condition that draws upon a wide range of threat-based data sources (Figure 33). We suggested the development of an integrated habitat condition assessment system comprising two parts – a) an initial "one-off" assessment using remote sensing data with sufficient rigour to inform the regional assessment process through the biodiversity significance model; b) an enhanced model, potentially with temporal dynamics, improved underpinnings, integrating different remote sensing data at highest possible resolution, with field verification and strategic monitoring to iteratively inform cumulative impact assessment. BHPBIO's own work in assessing vegetation condition and local infrastructure impacts on biodiversity will make an important contribution to this process.

6.1.2 BIODIVERSITY SIGNIFICANCE ASSESSMENT

We have demonstrated how disparate data sources and models can be integrated and applied to assess regional and local levels of Biodiversity Significance to inform decision making processes. Although not reported here in a regional and subregional context beyond the presentation of maps and data (via the

CSIRO Data Access Portal); the results can be overlain with other feature values such as species conservation significance, and areas of current and future potential impact including areas outside of BHPBIO's interest area. Given the regional nature of this assessment, and the importance of a shared approach to data to inform planning, the consideration of other areas of potential development will ensure consistency in decision making.

The underpinning datasets and models of species richness and local habitat condition need refinement before they can be considered sufficiently reliable to inform decision making. Our results therefore conservatively focus attention on the estimate of Biodiversity Significance representing natural uniqueness derived from the compositional turnover models only as the group average (Figure 35; individual outputs for each biological group are given in Appendix F) and Biodiversity Significance including regional condition (but excluding local condition and richness) (Figure 36). We consider the regional condition outputs to be robust because they consider only the overall (average) level of degradation across different environments (i.e., cells predicted to be compositionally similar as a result of scaling the environmental space by the relevant GDM) within the region, and not condition at any particular site of interest.

Relative to the regional composition of biodiversity across the Pilbara, areas demonstrating higher levels of natural uniqueness (assuming pristine habitat condition) include the coastal region, the riverine parts of the landscape, and locations throughout the Hamersley subregion (Figure 35). When regional condition estimates are included, accounting for overall levels of habitat degradation, the picture changes dramatically (e.g., Figure 36). The coastal and riverine areas of Biodiversity Significance are further emphasised and additional areas associated with these, and regionally more areas overall, have higher significance. The inclusion also of local habitat condition effects (Figure 37) moderates the pattern with a clearer focus on areas of significance and away from areas that are highly degraded. While some of this makes sense locally, it assumes a level of site-specific reliability in the mapping that is clearly questionable.

The demonstration of biodiversity significance with the inclusion of regional and local habitat condition provides an indication of how these models can be used to assess the cumulative impacts of development, or explore scenarios of proposed alternative areas for development to evaluate the marginal change in biodiversity significance by looking at each grid cell in turn, or a set of actions in grid cells. The biodiversity significance model can be used as the basis for these assessments provided sufficient rigour exists in the input models and data, coupled with a process of verification for critical land development decisions.

6.2 Published datasets

The dataset outputs (listed in Table 15) have been documented and made available to BHPBIO via CSIRO's data access portal (www.data.csiro.au). Access is limited to BHPBIO and CSIRO staff on the project.

Table 15. Published datasets provided to BHPBIO via CSIRO's data access portal (www.data.csiro.au).

PUBLISHED DATASET NAME	BRIEF DESCRIPTION	CITATION AND URL
Mean percent bare ground cover for the Pilbara region (2000-2011)	A satellite-based estimate of mean bare ground percent was derived from fractional cover Landsat TM bare ground grids at a 1 sec (30m) spatial resolution (Appendix E).	Perry J, Williams KJ, Ferrier S, Harwood T (2013) *Mean percent bare ground cover for the Pilbara region (2000-2011)*. v1. CSIRO. Data Collection. https://www.data.csiro.au/dap/landingpage?pid=csiro:7653
Interim biodiversity habitat condition layer, Pilbara Region	An interim biodiversity habitat condition layer was derived from best available data sets for land use, tenure and infrastructure and assigned scores based on potential anthropogenic disturbance (Section 4).	Perry J, Williams KJ, Ferrier S, Harwood T (2013) *Interim biodiversity habitat condition layer, Pilbara Region*. v1. CSIRO. Data Collection. https://www.data.csiro.au/dap/landingpage?pid=csiro:7652
Predicted richness and standard errors for five biological groups across the Pilbara	Species richness patterns for five biological groups (mammals, birds, reptiles, invertebrates – comprising ants, beetles and ground spiders – and vascular plants) were derived using Generalised Additive Modelling (Section 3.2 and Appendix D).	Williams KJ, Liu Y, Henderson B, Ferrier S, Perkins G, Harwood T (2013) *Predicted richness and standard errors for five biological groups across the Pilbara*. v1. CSIRO. Data Collection. https://www.data.csiro.au/dap/landingpage?pid=csiro:7635

PUBLISHED DATASET NAME	BRIEF DESCRIPTION	CITATION AND URL
GDM-scaled environmental predictors for compositional turnover in five biological groups across the Pilbara	Compositional turnover patterns among species assemblages in five biological groups (mammals, birds, reptiles, invertebrates – comprising ants, beetles and ground spiders – and vascular plants) were derived using Generalised Dissimilarity Modelling (Section 3.1 and Appendix C).	Williams KJ, Ferrier S, Perkins G, Manion G, Harwood T, Perry J (2013) *GDM-scaled environmental predictors for compositional turnover in five biological groups across the Pilbara*. v1. CSIRO. Data Collection. https://www.data.csiro.au/dap/landingpage?pid=csiro:7634
Biodiversity significance analyses and associated uncertainty for five biological groups across the Pilbara provided as an ESRI Map Package	Biodiversity significance analyses and associated uncertainty for five biological groups (mammals, birds, reptiles, invertebrates and vascular plants) and an aggregate of those groups. Six variants of the measure were derived within and without species richness and with or without habitat condition. Data are provided in an ArcMap10.1 Map Package which contains the inputs and outputs of Biodiversity Significance Analysis, includes custom legends and uncertainty overlays for model error (Section 5 and Appendix F).	Harwood T, Ferrier S, Williams KJ, Liu L, Perry J, Perkins G (2013): *Biodiversity significance analyses and associated uncertainty for five biological groups across the Pilbara provided as an ESRI Map Package*. v1. CSIRO. Data Collection. https://www.data.csiro.au/dap/landingpage?pid=csiro:7591

6.3 Future research and development

The work presented in this report (Sections 2 to 5) sets the scene for a potential second phase of work developing and applying new techniques for projecting cumulative impacts of development and conservation actions on biodiversity persistence throughout the Pilbara Bioregion, by implementing all the elements outlined in Figure 42.

The research and development framework outlined in Figure 42, aims to provide a rigorous, and defensible, means of assessing expected impacts of development and conservation activities on persistence of overall plant and animal diversity at a whole-bioregion scale by integrating, and significantly advancing, state-of-the-art techniques in environmental mapping, biodiversity modelling, and scenario analysis. The resulting capacity to estimate the level of persistence expected as a consequence of any spatially-explicit combination of development and conservation (protection and/or restoration) actions will, in turn, provide the foundation for three major forms of assessment:

- Mapping of the relative reduction (marginal loss) in whole-bioregion biodiversity persistence expected if a given type of development action were applied, in turn, to each and every location (fine-scaled grid cell) within the region. Conversely, mapping of the relative improvement (marginal gain) in persistence expected if a given type of conservation action were applied to each location in turn. Such maps would assist in identifying parts of the region where development is likely to have most, or least, impact on biodiversity persistence and, conversely, where conservation investment is likely to result in most, or least, benefit.

- Evaluation of the net impact that any spatially-explicit set of proposed (hypothetical) development and conservation actions is expected to have on whole-bioregion biodiversity persistence. This capability would provide an objective means of assessing biodiversity implications of alternative offsetting options, or whole alternative plans involving multiple development and conservation actions.

- Evaluation of the cumulative impact that development and conservation actions implemented over time are expected to be having on whole-bioregion biodiversity persistence in the longer term. This would provide a cost-effective means of tracking, and reporting on, cumulative impacts in terms of expected biodiversity outcomes – a capability that ideally needs to be coupled with long-term field

monitoring of actual outcomes to facilitate adaptive calibration and refinement of the underpinning models.
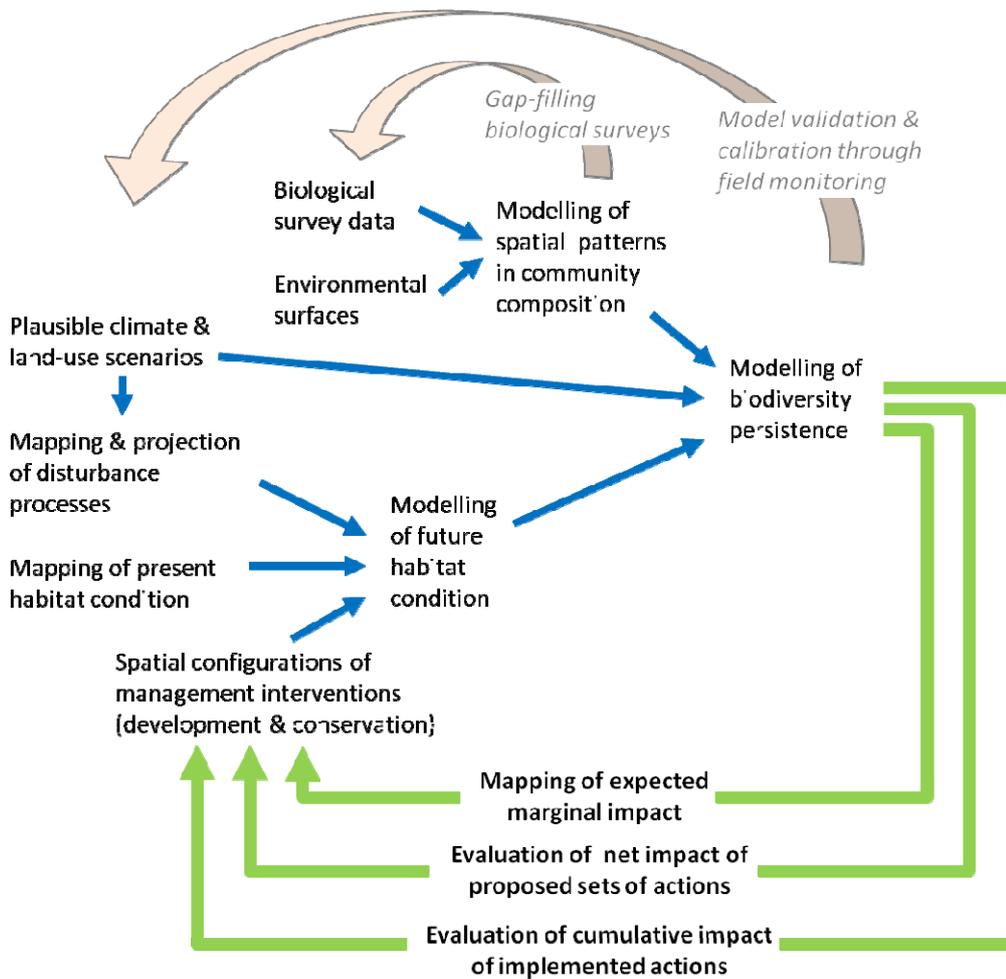
Figure 42. Proposed major components of an R&D framework for projecting cumulative impacts of development and conservation actions on biodiversity persistence.

A number of advances in community-level biodiversity modelling (all pursued by CSIRO in recent years) could serve as the starting point for developing this methodology, including:

- A general conceptual framework for whole-landscape modelling of biodiversity persistence, described by Ferrier and Drielsma (2010).

- Statistical approaches to modelling spatial patterns in biodiversity composition – e.g. Ferrier *et al.* (2007).

- Semi-mechanistic macroecological approaches to projecting impacts of climate and land-use change on biodiversity persistence – e.g. Mokany *et al.* (2012).

Application of such a research and development program would inject a whole new level of rigour and innovation into this work, across multiple fronts, all of which should have significant flow-on benefits for any subsequent assessment and/or decision underpinned by this methodology:

- Extended techniques for deriving fine-scaled environmental surfaces (particularly for substrate and microclimate) through integration of advances in remote sensing, terrain analysis, etc.

- New remote-sensing approaches to mapping habitat condition (building, in part, on some of the preliminary work described in Section 4 and Appendix E of this report).

- Richer consideration of the effects of multiple ecological (disturbance) processes of relevance in the Pilbara – including grazing, fire, invasive species, hydrology, etc.

- More rigorous consideration of the temporal dimension of impacts of development and conservation actions (time lags, state-and-transition dynamics, etc).

- Refinement of condition measures, and projections, to address differences in responses to disturbance processes between biological groups.

- Incorporation of potential interactions between biological groups.

- Capacity to utilise data from emerging metagenomic and phylogeographic sampling and analysis techniques.

- Development of rigorous strategies and procedures for model validation and calibration through field monitoring.

6.4 Key findings and recommendations

1. The data compiled by BHP for environmental assessment is a rich source of information, albeit often restricted to tenement areas. We focussed on certain aspects of these data for the analyses presented here; specifically, spatially referenced observations with taxonomic identifications at least to the species level that could be assigned an accepted name based on National censuses of plant and animal species. Two data sets presented particular limitations. Firstly, the invertebrate collections were mainly identified at the generic level or from poorly known groups based on targeted sampling, preventing their merger with the comprehensive surveys of beetles, ants and spiders across the Pilbara by the WA Department of Parks and Wildlife. Secondly, the comprehensive surveys of vascular plants across the Pilbara by the WA Department of Parks and Wildlife, which would help moderate the sampling bias associated with the surveys conducted by BHPBIO and their consultants, were unavailable at the time of this assessment. Finally, a further limitation of the BHPBIO data was the inability to easily distinguish the comprehensively surveyed sites for a particular biological group as reference sites for biodiversity modelling (richness and compositional dissimilarity). Different observation sources can be merged so long as covariates describing their similarities and differences can be generated and include in the modelling, to essentially weight their importance. This requires more detail about each survey's methodology and purpose, generally contained in survey reports.

2. The data underpinning the biodiversity significance models presented here can be enhanced through a concerted effort to compile all available surveys from among development interest groups across the region. Rather than a once-off compilation, we suggest data provider agreements with the Atlas of Living Australia and/or the Terrestrial Ecosystem Research Network who have established the protocols for aggregation and public use, or other appropriate entity to efficiently mediate data federation. Researchers, industry and consultants alike can then access the data, add value or assess the most cost-efficient requirements for filling knowledge gaps (e.g., through further survey) for a given purpose.

3. Species richness is a critical determinant of Biodiversity Significance, although secondary to models of compositional dissimilarity. The limitations of the current GAM models can be addressed through a more thorough examination of the available data sources and consideration of different statistical approaches. However, there is presently no specific best practice statistical approach and before investing in further analyses, gaps in the underpinning data require improvement through a more thorough examination of available sources as suitable 'reference sites' for modelling, and a targeted program of comprehensive surveys to fill gaps in environmental and taxonomic coverage of key indicator groups for conservation assessment.

4. Reliable assessment and monitoring of biodiversity habitat condition is also a critical determinant of Biodiversity Significance; and especially for any assessment of mining impacts and offsets. We present a novel framework toward solving this problem (Donohue *et al.* 2013) and outline a process for trialling this in the Pilbara (page 51). The successful application of this modelling framework, along with a network of monitoring sites to support calibration, will help integrate *in situ* and remote sensing data enabling a more reliable assessment of habitat condition applied at both regional and local scales for continuous reporting on the status of biodiversity and its use in scenario analyses.

5. Given current uncertainties in the assessment of species richness and site level habitat condition (outlined above), we can only recommend the use of Biodiversity Significance based on Equation 1 (natural uniqueness - Figure 35) and that incorporating regional condition (Equation 2 - Figure 36) for the current regional assessment process.

6. We recommend further interaction regarding the use of these outputs in BHPBIO's current and future strategic assessment work, to better resolve precisely which output to use for which purpose, or whether other variants of the measures presented here might be more applicable. For example, the map legends were designed for comparison within a group (with or without richness models included in the calculation), a different legend may be needed if the maps (e.g., Figure 35 and Figure 36) are used in isolation to ensure the maximum amount of information is conveyed.

7. Limitations in the biodiversity assessments identified here can be addressed through a program of research and development linked with BHPBIO's future planning and assessment needs, such as outlined in Figure 42.

Appendix A Distribution of biological data

This appendix presents maps of the distribution of biological data for different groups provided by BHPBIO and DPaW or downloaded from the Atlas of Living Australia (ALA). Numbers represent the number of records or plotted points for each map.

Araneae (spiders)
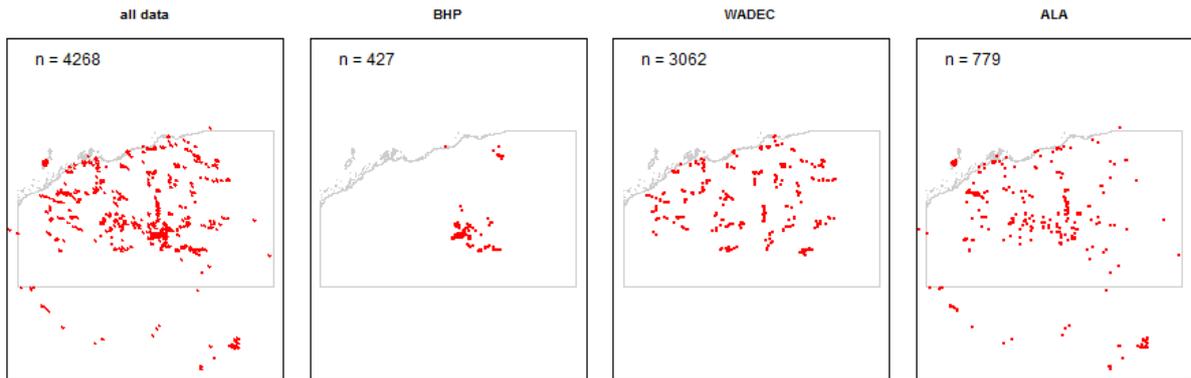
Aves (birds)
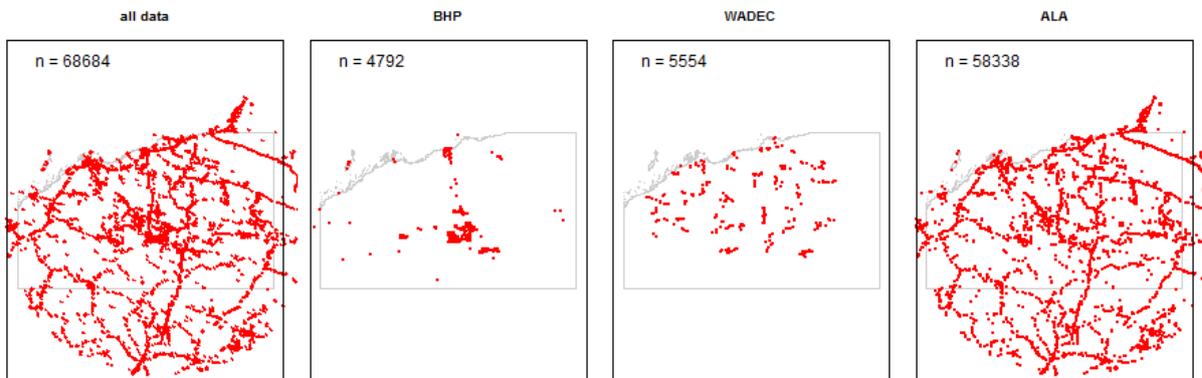
Carabidae (ground beetles)
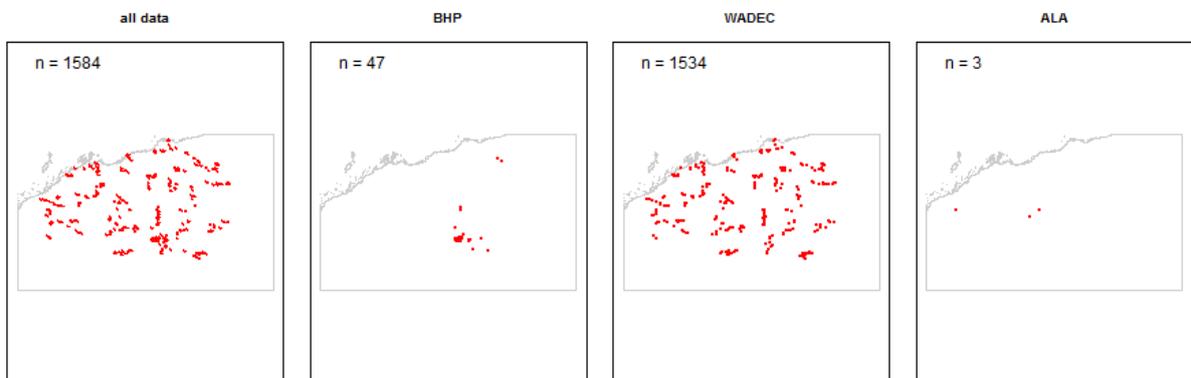
Coleoptera (beetles)
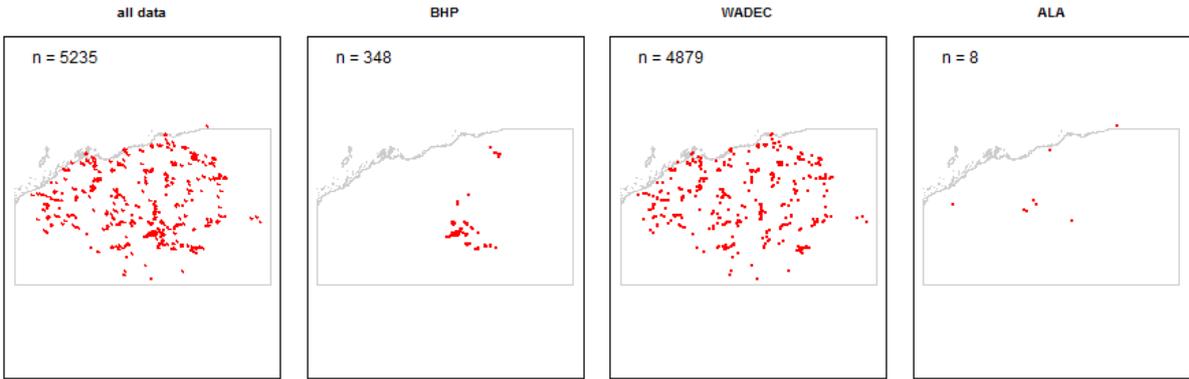
Diptera (flies)
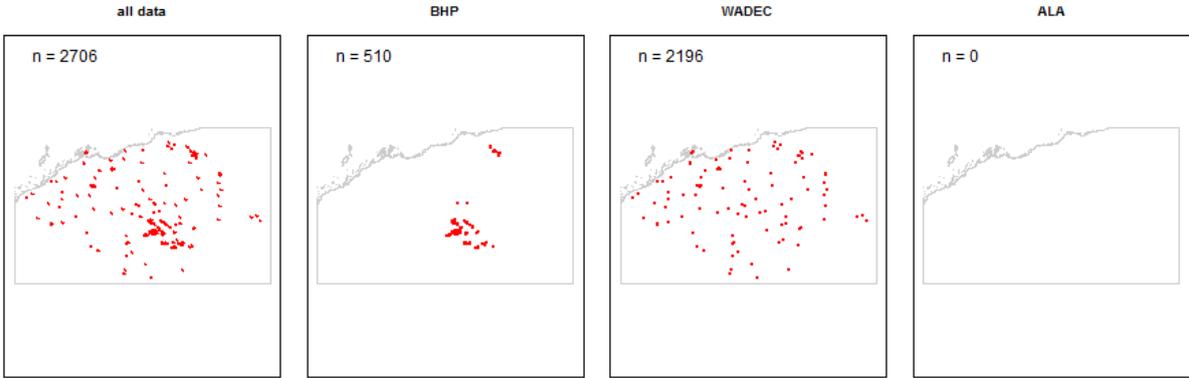
Formicidae (ants)
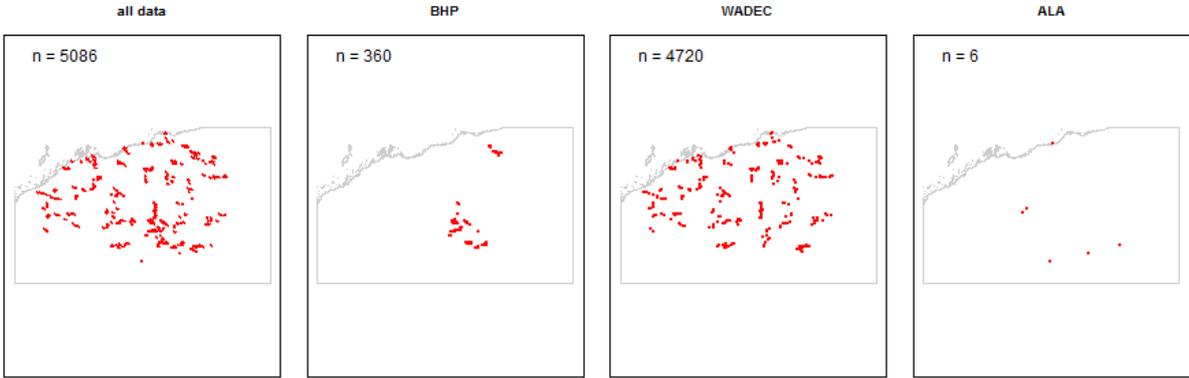
Gastropoda (snails)

Hemiptera (bugs)

all data	BHP	WADEC	ALA
n = 1172	n = 391	n = 767	n = 14
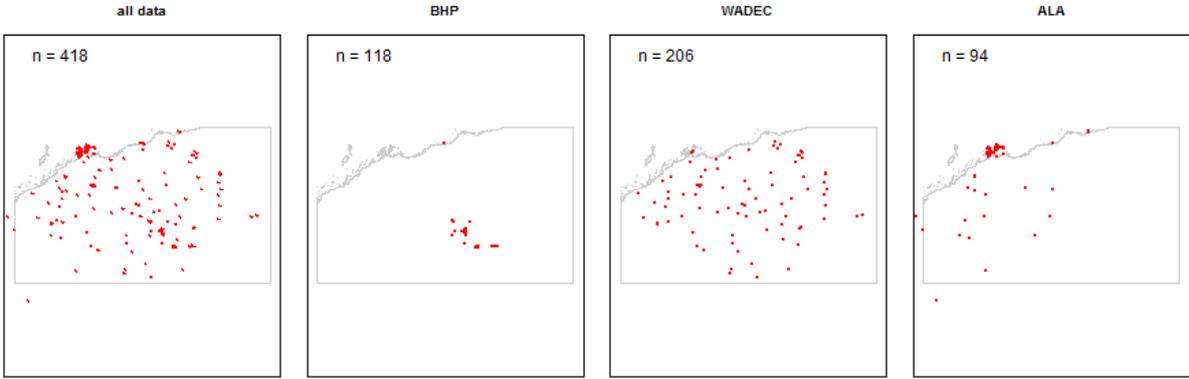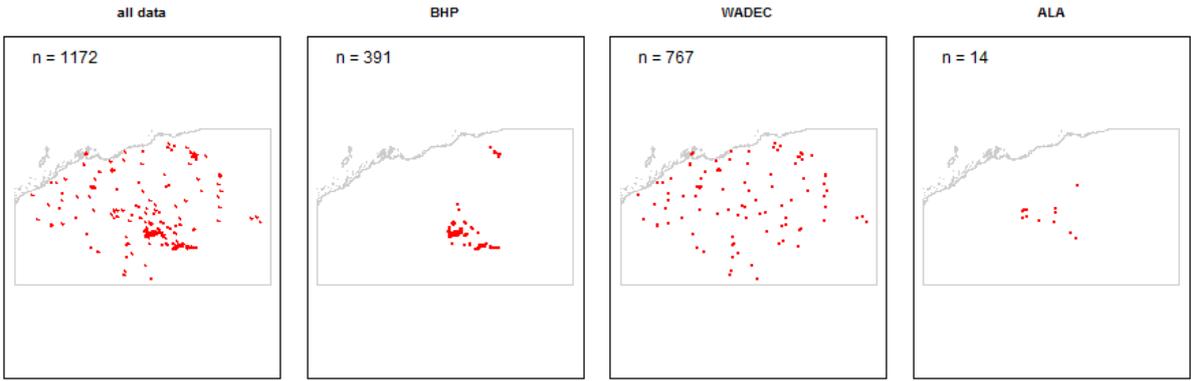
Hymenoptera (Ants, Bees, Wasps, Sawflies)

all data	BHP	WADEC	ALA
n = 5154	n = 410	n = 4720	n = 24

Mammalia (mammals)

all data	BHP	WADEC	ALA
n = 6283	n = 3418	n = 1031	n = 1834

Odonata (dragonflies)

all data	BHP	WADEC	ALA
n = 762	n = 1	n = 761	n = 0
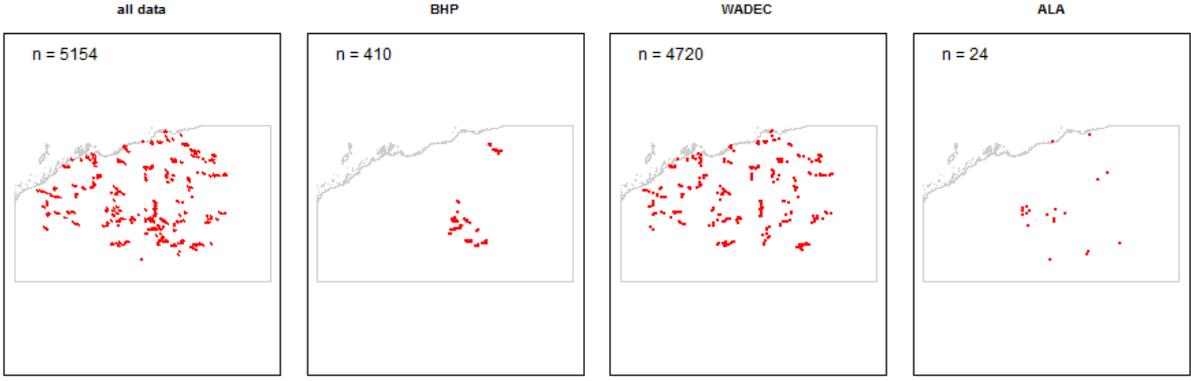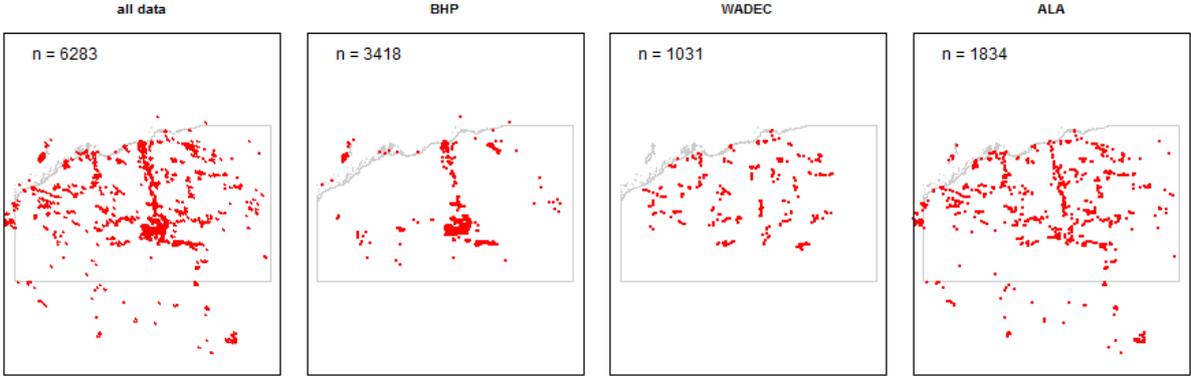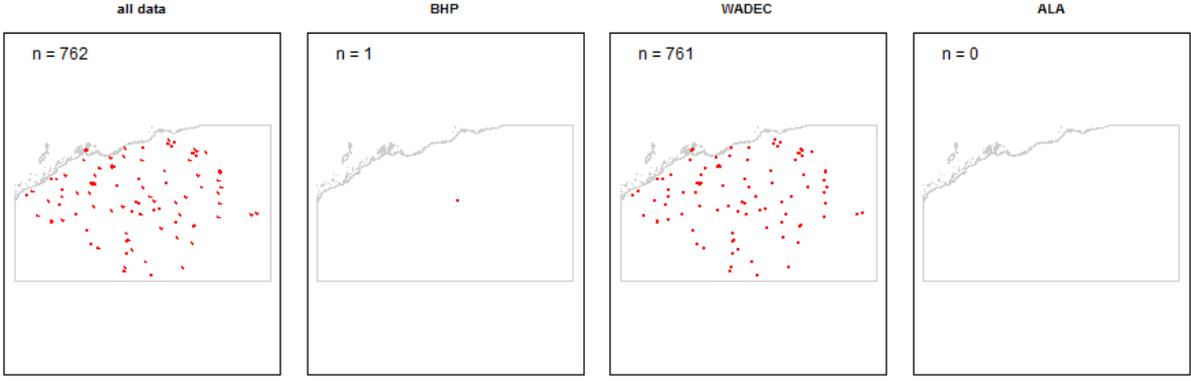
Reptilia (reptiles)
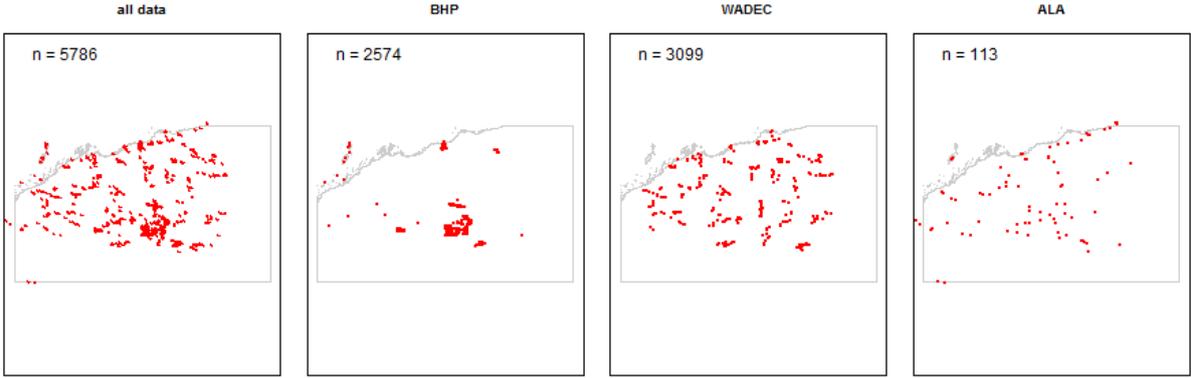
Scorpions
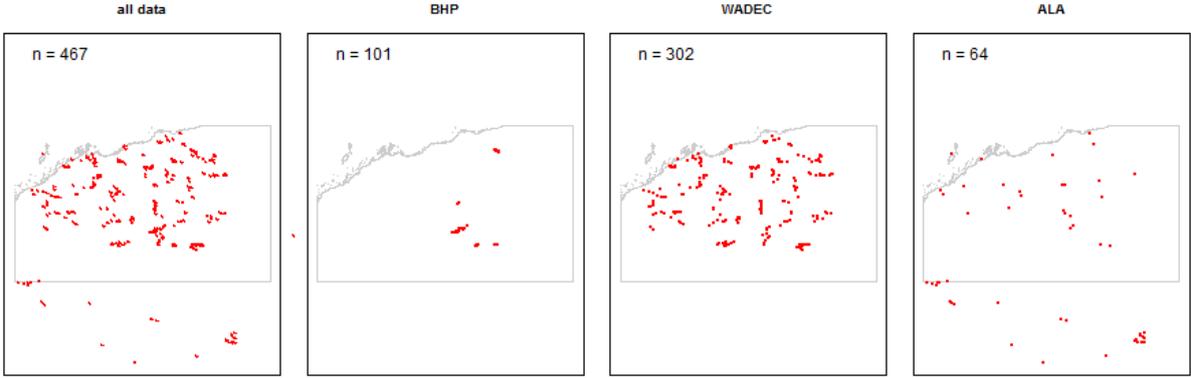
Vascular Plants
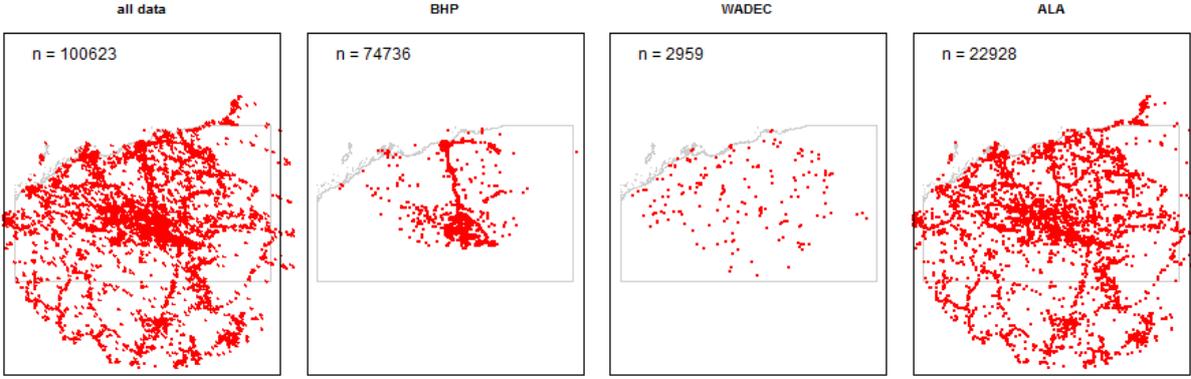
Appendix B Spatial environmental variables

The best available spatial environmental data were collated to represent a wide range of direct and indirect indicators of climate, regolith, landform, hydrology and land cover at 9-second grid resolution. Ordinal or continuous variables with consistent spatial dimensions are required for analysis using generalized dissimilarity modelling. Minor differences in spatial extent and data/no-data areas were filled using focal-mean or focal-majority algorithms for continuous and ordinal variables respectively, using ArcGIS software (ESRI 2011).

B.1   Climate

Long-term (30 years, 1975-2005) monthly variation in climate was derived from ANUCLIM v6.1 (Xu & Hutchinson 2011) using version 3 of the 9-second digital elevation model (DEM) for Australia (Hutchinson et al. 2008). Indices of annual and seasonal variation in a range of climatic variables were subsequently derived (Table 16). Vapour pressure deficit was derived using monthly dew point temperature, and wet and dry bulb temperatures based on the equations outlined in Allen et al. (1998). The psychrometric data method was used to estimate actual vapour pressure from gridded values of atmospheric pressure based on altitude from the DEM. For relative humidity we adopted the equation used by the Australian Bureau of Meteorology (after Abbott & Tabony 1985).

These climatic variables characterise general patterns of seasonal wetting and drying but lack detail about inter-annual variability. This, combined with the sparse distribution of weather stations throughout the Pilbara, limits the accuracy of the climatic signal and its correlation with patterns of regional variation in the biota. For this reason, two indirect location-based predictors were included as proxies for unexplained variation correlated with climate – Euclidean distance from coast and elevation. McKenzie et al. (2009) also used these indirect variables in their analyses of Pilbara biodiversity patterns (e.g., Gibson & McKenzie 2009). The elevation data were sourced from the 3 second DEM for Australia which is a derivative of the 1 second DEM (Gallant & Read 2009; Gallant 2011; Geoscience Australia & CSIRO Land & Water 2011). The elevation mean and range within each 9-second grid were compiled as candidates for explaining regional to local elevation heterogeneity.

Table 16. Indices of annual and seasonal variation in climate compiled for the Pilbara analysis region.

Code	Description
ADEFIE	Maximum monthly precipitation deficit (Rain-Evap) (mm)
ADEFXE	Minimum monthly precipitation deficit (Rain-Evap) (mm)
EVAPIE	Minimum monthly evaporation (mm)
EVAPXE	Maximum monthly evaporation (mm)
RAINIE	Precipitation of the driest month (mm)
RAINXE	Precipitation of the wettest month (mm)
RADNIE	Minimum monthly rainfall-modified solar radiation (MJ/m^2/day) (flat-surfaces)
RS06F	Solar radiation in June modified by topography
RS12F	Solar radiation in December modified by topography
RADNXE	Maximum monthly rainfall-modified solar radiation (MJ/m^2/day) (flat-surfaces)
MINTIE	Minimum temperature coldest month (°C)
MINTXE	Minimum temperature warmest month (°C)
MAXTXE	Maximum temperature coolest month (°C)
MAXTIE	Maximum temperature hottest month (°C)

Code	Description
TRNGIE	Minimum monthly diurnal temperature range (°C)
TRNGXE	Maximum monthly diurnal temperature range (°C)
RH2MAXE	Maximum monthly relative humidity (%)
RH2MINE	Minimum monthly relative humidity (%)
VPD_MAXE	Maximum monthly vapour pressure deficit (KPa)
VPD_MINE	Minimum monthly mean vapour pressure deficit (KPa)
BIO03E	Isothermality 2/7
BIO07E	Temperature Annual Range (5-6)
BIO29E	Highest Period Moisture Index
BIO30E	Lowest Period Moisture Index
BIO31E	Moisture Index Seasonality (C of V)
DISTCOAST	Euclidean distance to coast in degrees based on 9-second grid
ELEVATIONME	mean elevation within 9-second grid based on 3-second DEM

B.2   Regolith

Because spatial data describing soils is limited in resolution, a wide range of candidate variables that are potentially correlated with patterns in soil variability at the site level were compiled. These variables fall into four classes: 1) attributes derived from soil maps; 2) models of soil properties; 3) attributes derived from geology maps; and 4) geophysics variables such as magnetic, gravity and radiometric data (Table 17).

Two sources of attributes from soil maps were available to the study; one set based on an interpretation of the Atlas of Australian soils by McKenzie and coworkers (McKenzie & Hook 1992; McKenzie *et al.* 2000a; Western & McKenzie 2004), and another set based on a composite of best available soil maps (McKenzie et al. 2005; Jacquier 2011d, b, c, a), available through the Australian Soil Resource Information System (http://www.asris.csiro.au/themes/NationalGrids.html). Where these attributes represent the same property, the more recent set based on the best composite of maps was used, even though map boundaries may influence the spatial appearance of the prediction.

Variation in surface geology for Australia has been compiled by Geoscience Australia from 1:250,000 source data (Liu et al. 2006; Raymond et al. 2007a; Raymond et al. 2007b; Raymond et al. 2007c; Whitaker et al. 2007; Stewart et al. 2008; Whitaker et al. 2008). However, only two attributes are currently available in continuous or ordinal form: inherent rock fertility (De Vries 2009) and geological age based on the timescales database (Laurie et al. 2008) compiled by Williams and co-workers (Williams *et al.* 2010b, a; Williams *et al.* 2012) at 0.01 resolution.

Digital soil mapping is a rapidly developing field of science (McBratney et al. 2003) and a number of modelled soil properties are becoming available nationally with potential application in ecology (e.g., Viscarra-Rossel et al. 2010a; Viscarra-Rossel et al. 2010b; Viscarra-Rossel & Chen 2011; Viscarra Rossel 2011; Gray et al. 2012; Wilford 2012). We explore the potential predictive utility of several of these modelled soil properties: weathering intensity index (Wilford 2012); three principal components of soil colours (Viscarra-Rossel et al. 2010b) and three clay minerals (Viscarra Rossel 2011) at two soil depth profiles – 0-20cm and 60-80 cm.

Geophysics data is another potential source of information about patterns of variability in the soil environment or an indication of surface geology features that may influence soil formation. National coverage exists for gravity (Geoscience Australia 2009) and magnetic (Milligan 2010b, a) anomalies to support exploration geology. Gravity anomalies in units of acceleration show the effects of different rock densities in the subsurface compared with the surrounding earth's mantle (Gunn 1997). Mountains are usually negative because the rock density at the base is lower relative to surrounding areas. Local positive anomalies may indicate metallic ores. Anomalies can help to distinguish sedimentary basins whose fill

differs in density from that of the surrounding region. The 2009 edition of the gravity grid of Australia and surrounding areas is a compilation of over 1.4 Million onshore gravity stations and offshore marine gravity data derived from satellite altimetry with a cell size of 800m for the Lambert Conic Conformal Projection grids and 0.00833333 degrees for the geodetic grids (Geoscience Australia 2009; Percival 2010).

Magnetic anomalies also provide insights into the distribution of magnetically susceptible minerals within the Earth's crust. The source of magnetization of rocks is primarily from magnetic induction by the earth's field of particles of magnetite within the rocks. The lighter colored rocks such as granite and porphyries have much less magnetite than the dark basic rocks such as gabbro and diabase (Nettleton 1971). Magnetic minerals become concentrated in zones which highlight the structure of the crust and magnetic anomalies can detect magnetic signatures below surface cover (regolith and sedimentary basins). Magnetism may also be affected by weathering and sedimentation processes, as well as rock metamorphism and hydrothermal reactions (Clark 1997). Because magnetism captures information about different rock types, it may also reflect disjunctions in regolith type. The fifth edition total magnetic intensity grid of Australia has been derived with a grid cell spacing of 3-arc seconds (Milligan 2010a; Percival 2010).

Radiometric or gamma-ray spectrometry data is another potential source of information about the structure and composition of the top 30-40cm of the land surface. All rocks and soils contain radioactive isotopes, and almost all the gamma-rays detected near the Earth's surface are the result of the natural radioactive decay of potassium, uranium and thorium. Changes in lithology, or soil type, are often accompanied by changes in the concentrations of radioelements and may also indicate mineral deposits. Potassium, uranium and thorium behave quite differently from one another during weathering and pedogenesis and in combination with digital terrain data can be used to define regolith properties such as weathering intensity (Wilford 2012). Quantitative soil characterisation based on radiometric data is an area of continuing research (Beckett 2003, 2007). We obtained the 2010 edition of the radiometric map of Australia dataset (Geoscience Australia 2010). This dataset comprises grids of potassium (K), uranium (U) and thorium (Th) element concentrations, and derivatives of these grids, that were derived by seamlessly merging over 550 airborne gamma-ray spectrometric surveys in the national radioelement database (Percival 2010) using the method described by Minty et al. (2009). The original survey grids were levelled and then re-sampled, using minimum curvature (Briggs 1974), onto the Radiometric Map of Australia Grids with a cell size of about 100m (0.001 degrees).

Table 17: Indices of regolith (regolith) variability compiled for the Pilbara analysis region.

Group	Code	Description
Composite soil map	BD30E	bulk density to 30cm derived from ASRIS 9-second gridded composite soil map variables
	CLAY30E	percent soil clay content to 30cm derived from ASRIS 9-second gridded composite soil map variables
	PAWC1ME	Plant available soil water content (mm/100cm) derived from ASRIS 9-second gridded composite soil map variables
Atlas of Australian soils	SOLDEPTHNE	solum depth in metres based on a weighted average of PPF mosaics
	NUTRIENTSNE	Gross soil nutrient status based on weighted average of PPF mosaics
	COARSENE	percentage of coarse fragments throughout the soil profile based on weighted average of PPF mosaics
	HSTRUCTNE	Hydrological scores for grades of pedality based on correlations with measured steady infiltration rates for a wide range of soils, as determined by Lin *et al.* (1999): single grain (50), massive (0), weak (1), moderate (5), or strong (25) based on weighted average of PPF mosaics
	KSATNE	solum average of the median horizon saturated hydraulic conductivity weighted by the depth of each horizon and based on a weighted average of PPF mosaics
	CALCRETENE	The presence (1) or absence (0) of calcrete in or below the soil profile based on weighted average of PPF mosaics
Terrain-soil index	CTIDEPTHU2E	The soil depth in metres was derived from a topographic wetness index (twifd80_011e) scaled by map unit soil depth range (Claridge *et al.* 2000), derived from

Group	Code	Description
		attributes for the principal soil profile form developed for the Atlas of Australian Soils (McKenzie *et al.* 2000a). Specifically, the minimum, mode and maximum values of the FD8 CTI based on 9-second DEM v3.1 were scaled by minimum, mean and maximum soil depth values respectively applied to each soil map unit (across multiple polygons).
	CTIPAWCU2E	The soil water holding capacity in mm was derived from a topographic wetness index (twifd80_011e) scaled by map unit soil depth range (Claridge *et al.* 2000), derived from attributes for the principal soil profile form developed for the Atlas of Australian Soils (McKenzie *et al.* 2000a), and then multiplied by the soil water holding capacity in one metre based on the ASRIS 250m best-composite soil map data (Jacquier 2011a).
Modeled clay minerals	ILL20ME	9-second grid zonal mean of % abundance modelled illite clay minerals in surface soil (0-20cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
	ILL80ME	9-second grid zonal mean of % abundance modelled illite clay minerals in subsurface soil (60-80cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
	KA020ME	9-second grid zonal mean of % abundance modelled kaolinite clay minerals in surface soil (0-20cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
	KA080ME	9-second grid zonal mean of % abundance modelled kaolinite clay minerals in subsurface soil (60-80cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
	SME20ME	9-second grid zonal mean of % abundance modelled smectite clay minerals in surface soil (0-20cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
	SME80ME	9-second grid zonal mean of % abundance modelled smectite clay minerals in subsurface soil (60-80cm) derived from 3-second grid source data using the FOCALMEAN function with a 3x3 cell window, and expanded to fill null values within pilmask1
Modeled soil spectra	PC1ME	Spectra of surficial soils – Principal component 1 (3-second grid resolution, approx 90m) (Raphael Viscarra Rossel) - zonal mean in 9-second grid
	PC2ME	Spectra of surficial soils – Principal component 2 (3-second grid resolution, approx 90m) (Raphael Viscarra Rossel) - zonal mean in 9-second grid
	PC3ME	Spectra of surficial soils – Principal component 3 (3-second grid resolution, approx 90m) (Raphael Viscarra Rossel) - zonal mean in 9-second grid
Geology	MINFERTFE	An index of inherent rock fertility for the 1:1 million geology of Australia that ranks lithological types on a 1 to 6 scale from rocks that are extremely siliceous (>90% silica) with an extremely low base content (<3% Ca, Mg, Fe oxides) to those that are ultramafic (<45% silica and >30% base content)
	GEOLRNGAGGNE	Range in geological age in millions of years before present derived from the 1:1 million surface geology of Australia
	GEOLMNAGGNE	Mean geological age in millions of years before present derived from the 1:1 million surface geology of Australia
Modeled weathering	WII_OZ2ME	Weathering intensity index (Wilford 2012) - zonal mean sampled to 9-second from original 3-second grid
Magnetic	PILMAGME	Mean of magnetic anomalies based on 80m grid zonal statistics in 9-second grid
Gravity	PILGRAVME	Gravity anomalies resampled from 800m grid to 9-second grid
Radiometric data	DOSEME	Terrestrial dose rate derived as a linear combination of the filtered K, U and Th grids (see Minty et al., 2009), nG/h - zonal mean sampled to 9-second from 3-second grid
	TOTALDOSEME	Total dose rate due to terrestrial and cosmic radiation, derived by adding the estimated cosmic dose contribution to the filtered dose due to terrestrial sources described above, nG/h - zonal mean sampled to 9-second from 3-second grid
	PCTKME	Low-pass filtered K element concentrations, % K (potassium) - zonal mean sampled to

Group	Code	Description
		9-second from 3-second grid
	PPMTHME	Low-pass filtered Th element concentrations, ppm eTh (Thorium) - zonal mean sampled to 9-second from 3-second grid
	PPMUME	Low-pass filtered U element concentrations, ppm eU (Uranium) - zonal mean sampled to 9-second from 3-second grid
	RATIO_TKME	Ratio of Th over K derived from the filtered Th and K grids, dimensionless - zonal mean sampled to 9-second from 3-second grid
	RATIO_UKME	Ratio of U over K derived from the filtered U and K grids, dimensionless - zonal mean sampled to 9-second from 3-second grid
	RATIO_UTME	Ratio of U over Th derived from the filtered U and Th grids, dimensionless - zonal mean sampled to 9-second from 3-second grid
	RATIO_U2TME	Ratio of U2 over Th derived from the filtered U and Th grids, dimensionless - zonal mean sampled to 9-second from 3-second grid

B.3   Landform

Variation in topography and landform can be captured by indices derived from digital terrain models (Wilson & Gallant 2000a). Gallant and co-workers have been generating terrain indices based on the 1sec digital elevation model for Australia (Geoscience Australia & CSIRO Land & Water 2010). These indices vary in complexity from simple, such as slope (Gallant *et al.* 2011f), relief (Gallant *et al.* 2011i) and elevation diversity (Gallant et al. 2011a), to more complex algorithms such as topographic wetness indices (Gallant *et al.* 2011j) based on contributing area (Gallant *et al.* 2011e) and landform shape (Gallant *et al.* 2011g, h). Classes of topographic position commonly observed in the field (Speight 2009) also can be interpreted from a digital elevation model (Gallant & Austin 2012) or derived using multiresolution methods (Gallant & Dowling 2003). We compiled a few of the available indices describing landform diversity from existing datasets derived from different resolution digital elevation models (DEMs) and selected those that best represent landform at 9-sec resolution (Table 18). Two indices – topographic position index and associated binary mask for flat areas – are designed to complement each other because the topographic index does not accurately represent flat areas. Therefore both are included as candidate predictors in a model.

Terrain derivatives developed by Gallant and co workers (Gallant et al. 2011j, e, d, a, c, g, h, f, i, b; Gallant & Austin 2012) from the 3-second DEM-S (smoothed digital elevation model, Geoscience Australia 2013b) were aggregated to 9-second resolution using the median value for ordinal variables and the mean value for continuous variables (Table 18). Only a few of these are expected to be useful predictors.

Table 18. Indices of landform diversity compiled for the Pilbara analysis region.

CODE	Description
SLOPEDEG	mean slope in degrees within 9-second grid based on 3-second DEM (zero-filled null values)
MRRTFE	multi-resolution ridgetop flatness index based on 9-second DEM V3.1
MRVBFE3	Multiresolution valley bottom flatness index class is a topographic index designed to identify areas of deposited material at a range of scales based on the observations that valley bottoms are low and flat relative to their surroundings and that large valley bottoms are flatter than smaller ones. Zero values indicate erosional terrain with values 1 and larger indicating progressively larger areas of deposition. There is some evidence that MrVBF values correlate with depth of deposited material. The 3-second data were aggregated at 9-second using the raster grid median value.
PLANCURE3	Plan (or contour) curvature is the rate of change of aspect (orthogonal to the slope) and represents topographic convergence or divergence. It is significant for water movement across the landscape, i.e., the accumulation or dispersion of water. Plan curvature was calculated from 3-second DEM-S using the finite difference method (Wilson & Gallant 2000b). The different spacing in the E-W and N-S directions due to the geographic projection of the data was accounted for by using the actual spacing in metres of the grid points calculated from the latitude. The 3-second data were aggregated at 9-

CODE	Description
	second using the raster grid mean value.
PROFCURE3	Profile curvature is the rate of change of potential gradient down a flow line and represents the changes in flow velocity down a slope. It is significant for flow acceleration, erosion/deposition rates and geomorphology. Profile curvature was calculated from 3-second DEM-S using the finite difference method (Wilson & Gallant 2000b). The different spacing in the E-W and N-S directions due to the geographic projection of the data was accounted for by using the actual spacing in metres of the grid points calculated from the latitude. The 3-second data were aggregated at 9-second using the raster grid mean value.
SLPRELIEFE3	Slope relief landform pattern classification is based on Speight (2009). The slope_relief layer is an implementation of the classification of erosional landform patterns characterised by relief and modal slope as defined in Table 5 of Speight (2009). Modal slope has been replaced by median slope, since this is considered more amenable to automated processing, and the second highest relief class (90 - 300m) has been split into two classes, 90 - 150m and 150 - 300m; to connect the result with international work where the 150m relief threshold is used and due to the perceived rarity of relief over 150m in Australia. The 3-second data were aggregated at 9-second using the raster grid median value.
TPICLASSE3	Topographic position classification index identifies the upper, middle and lower parts of the landscape. A mask is also included that identifies where topographic position cannot be reliably derived in low relief areas. The 3-second data were aggregated at 9-second using the raster grid median value.
TPIMASKE3	Topographic position mask identifies relatively flat areas where the finite accuracy of a DEM limits its ability to discriminate topographic position. The mask included with the TPI layer identifies areas that are too flat to reliably identify upper, middle and lower landscape positions. It is based on the 'Slope-Relief' classification and the TPI mask has values of 1 where there is sufficient relief for TPI to be meaningful and 0 where TPI cannot be reliably used. The 3-second data were aggregated at 9-second using the raster grid median value.
CONAREAE3	The partial catchment contributing area in m^2 was computed using multiple flow directions on hillslopes and ANUDEM-derived flow directions in channels. The contributing area was computed on 1 degree tiles with 200 cell (about 5km) overlaps so the areas in channels do not account for catchments beyond that size (hence the use of PARTIAL in the name). The primary purpose of this product was to calculate topographic wetness index (TWI; Gallant and Wilson, 2000) for which full contributing areas in channels are not necessary. This product should not be used to represent contributing areas of catchments larger than 5km across. The 3-second data were aggregated at 9-second using the raster grid mean value.
TWIE3	The topographic wetness index is calculated as log_e(specific catchment area / slope) and estimates the relative wetness within a catchment. TWI was calculated from 3-second DEM-H following the methods described in Gallant and Wilson (2000). The program uses a slope-weighted multiple flow algorithm for flow accumulation, but uses the flow directions derived from the interpolation (ANUDEM) where they exist. Note that the partial contributing area product does not always represent contributing areas larger than about $25km^2$ because it was processed on overlapping tiles, not complete catchments. This only impacts on TWI values in river channels and does not affect values on the land around the river channels. Since the index is not intended for use in river channels this limitation has no impact on the utility of TWI for spatial modelling larger than 5km across. The 3-second data were aggregated at 9-second using the raster grid mean value.
SLPFM300E3	The 300m focal median of percent slope was derived from the 3-second DEM-S using a moving window of 300m. Slope measures the inclination of the land surface from the horizontal. Percent slope represents this inclination as the ratio of change in height to distance. The focal mean of percent slope can be used as a surrogate for modal slope in landform pattern analysis. The 3-second data were aggregated at 9-second using the raster grid mean value.
ELEVFR300E3	The elevation focal range (metres) within 300m was derived from the elevation 3-second DEM-S. The elevation range measures the full range of elevations within a circular window and can be used as a representation of local relief. Focal range using a 300m window was calculated for each grid point from DEM-S using a 300m kernel. The different spacing in the E-W and N-S directions due to the geographic projection of the data was accounted for by using the actual spacing in metres of the grid points, and recalculating the grid points included within the kernel extent for each 1° change in latitude. The 3-second data were aggregated at 9-second using the raster grid mean value.

CODE	Description
ELVFR1000E3	The elevation focal range (metres) within 1000m was derived from the elevation 3-second DEM-S. The elevation range measures the full range of elevations within a circular window and can be used as a representation of local relief. Focal range using a 1000 m window was calculated for each grid point from DEM-S using a 1000 m kernel. The different spacing in the E-W and N-S directions due to the geographic projection of the data was accounted for by using the actual spacing in metres of the grid points, and recalculating the grid points included within the kernel extent for each 1° change in latitude. The 3-second data were aggregated at 9-second using the raster grid mean value.
HILLSHADE	The hillshade dataset was derived from the 3-second DEM-S to representing general topographic shadowing effects

B.4 Vegetation and land cover

Several indices related to land cover and vegetation were derived as indicators of terrestrial habitat structure. Vegetation cover across the semi-arid climate of the Pilbara region is a good indicator of more productive habitats and areas in which some types of biodiversity may congregate. We therefore compiled or generated datasets as indicators of biotic response (Table 19).

An index of mean annual evapotranspiration was derived from the 10 year monthly data derived from MODIS by Guerschman *et al.* (2009). This variable indicates where the pooling of water in the landscape supports evapotranspiration associated with vegetation cover. We also compiled estimates of gross primary productivity from remote sensing of vegetation greenness partitioned into three functional plant types in the TMS scheme, a framework that links canopy leaf properties, vegetation structure and cover with resource availability (Berry & Roderick 2006; Berry *et al.* 2007; Mackey *et al.* 2012).

We further used an estimated of woody vegetation cover derived from Landsat remote sensing (Furby et al. 2009) for the years 2006 and 2009 (DCCEE 2012). Two indices were derived to represent the extent of woody vegetation within a grid cell and the nearest distance from any location to a vegetated grid cell.

Table 19. Indices of vegetation diversity and habitat compiled for the Pilbara region.

Code	Description
FT_NVEGE	Turgor (T) functional plant type derived from MODIS FPAR interception (9-second grids). Turgor leaves rely on the turgor pressure of the protoplasm for structural support. They are short-lived (e.g., leaves of deciduous trees and shrubs, and soft-leaved herbaceous species), have high photosynthetic rates per unit volume and low C:N ratios.
FM_NVEGE	Mesic (M) functional plant type derived from MODIS FPAR interception (9-second grids). Mesic leaves are longer-lived and may occur on trees, shrubs and herbaceous plants. The structural support for M leaves is provided by cell walls. M leaves have lower rates of photosynthesis per unit volume and higher C:N ratios than T leaves.
FS_NVEGE	Sclerophyll (S) functional plant type derived from MODIS FPAR interception (9-second grids). Sclerophyll leaves may occur on trees, shrubs and herbaceous plants and are longer-lived. The structural support for S leaves is provided by cell walls. S leaves have a smaller surface to volume ratio than M leaves and have the lowest photosynthetic rates and the highest C:N ratios.
MODISEA	Mean annual MODIS-derived potential evapotranspiration
EDISTFORST	The Euclidean distance to woody vegetation was derived from the aggregated forest extent dataset (DCCEE 2012) combining the 2006 and 2009 epochs. Forest is defined as vegetation with a minimum 20 per cent canopy cover, potentially reaching 2 metres high and a minimum area of 0.2 hectares. The EUCDISTANCE function in ARCINFO ESRI 2011) was applied in geographic distance units.
FORST06	The extent of woody vegetation in 2006 within 9-second grids derived from the aggregated forest extent dataset (DCCEE 2012). Forest is defined as vegetation with a minimum 20 per cent canopy cover, potentially reaching 2 metres high and a minimum area of 0.2 hectares.

B.5   Hydrologic

Additional spatial environmental indices were developed to describe the aquatic and riparian environments of the Pilbara region. Little attention has been given to the development of spatially explicit context indices describing aquatic habitats and their relationship to the surrounding landscape (although see Kennard et al. 2010; Turak et al. 2010).

We approached this by first developing an ecological rationale for why particular attributes of the environment may be important based on scientific understanding of aquatic systems in the Pilbara region. Indices subsequently derived from available data sources are described in Table 20. In addition to stream order defining the position of the catchment in the landscape and distance to water bodies of different types, a series of terrain attributes were compiled as a proxy for topographically relevant hydrological information (listed in Table 18). The Bureau of Meteorology's GEOFRABRIC database provided the source information for these variables (Bureau of Meteorology 2010).

Table 20. Customised indices for riparian and aquatic habitats in the Pilbara analysis region.

Code	Description
EDISTHYDRO	The Euclidean distance to water points dataset was derived using data combined from the GEOFABRIC 1.0 and WA Department of Water mapping sources (Pilbara_hydropoints.shp) (Bureau of Meteorology 2010; Department of Water 2011) and the EUCDISTANCE function in ARCINFO (ESRI 2011) in geographic distance units.
EDISTMAJOR	The Euclidean distance to major drainage lines was derived using data derived from GEOFABRIC 1.0 (drainage_lines_major.shp) (Bureau of Meteorology 2010) and the EUCDISTANCE function in ARCINFO (ESRI 2011) in geographic distance units.
EDISTMINOR	The Euclidean distance to minor drainage lines was derived using data derived from GEOFABRIC 1.0 (drainage_lines_minor.shp) (Bureau of Meteorology 2010) and the EUCDISTANCE function in ARCINFO (ESRI 2011) in geographic distance units.
EDISTPEREN	The Euclidean distance to perennial water bodies and drainage was derived using data derived from GEOFABRIC 1.0 (Bureau of Meteorology 2010) and the EUCDISTANCE function in ARCINFO (ESRI 2011) in geographic distance units.
ECLIFFLINE	The Euclidean distance to cliff lines was derived using the GEOFABRIC 1.0 AHGF SH_Cartographic geodatabase (extracted as AHGF_cliffline_pilbara.shp) (Bureau of Meteorology 2010) and the EUCDISTANCE function in ARCINFO (ESRI 2011) in geographic distance units. Cliff lines are also shown as terrain break lines in the GEOFRABRIC dataset.
SALINITY2F	The median water table salinity (mg/L) was derived from geofrabric AHGF_GW_Interim dataset (extracted as WaterTableSalinity_median_Pilbara.shp) (Bureau of Meteorology 2010).
LASTPFAFE	The Pfafstetter attribute representing the highest catchment stream order was derived from GEOFABRIC AHGFCatchment "lastpfaf" attribute (extracted as AHGFcatchment_lastpfaf_Pilbara.shp) (Bureau of Meteorology 2010).

Appendix C Generalised dissimilarity models, GDM
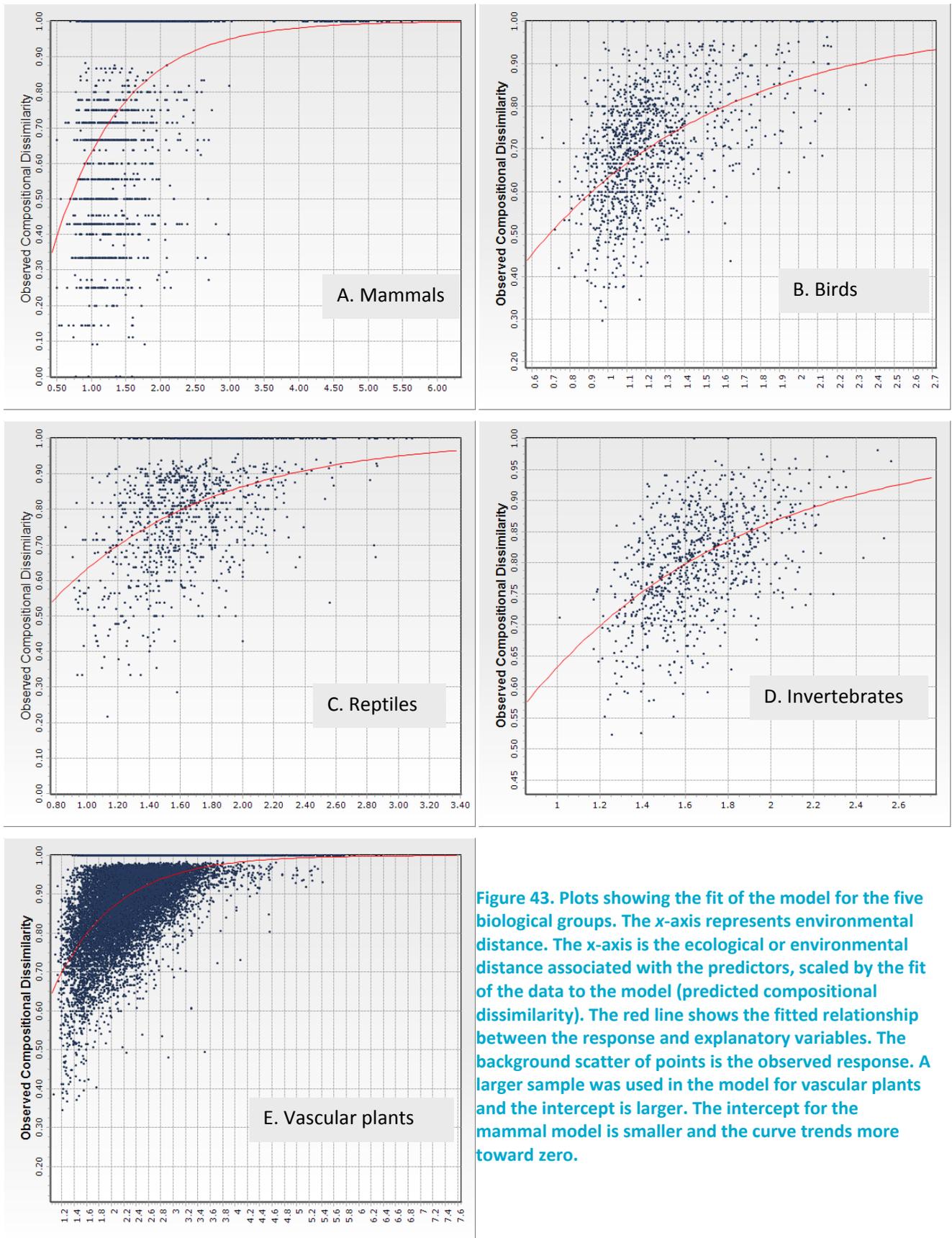
Figure 43. Plots showing the fit of the model for the five biological groups. The *x*-axis represents environmental distance. The x-axis is the ecological or environmental distance associated with the predictors, scaled by the fit of the data to the model (predicted compositional dissimilarity). The red line shows the fitted relationship between the response and explanatory variables. The background scatter of points is the observed response. A larger sample was used in the model for vascular plants and the intercept is larger. The intercept for the mammal model is smaller and the curve trends more toward zero.
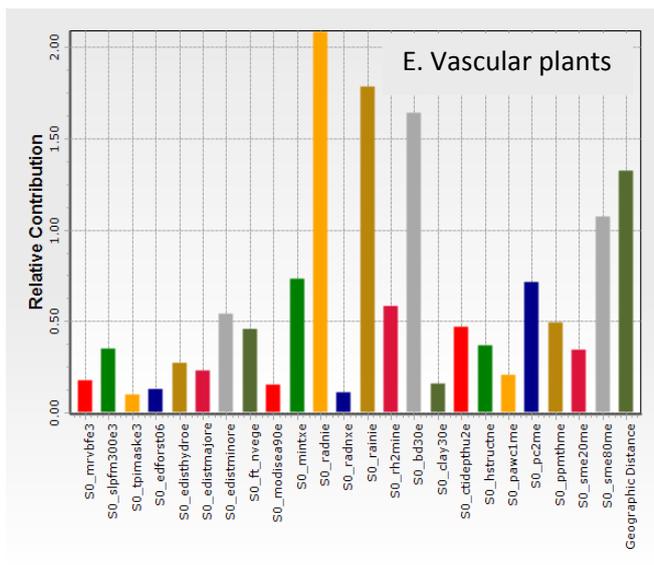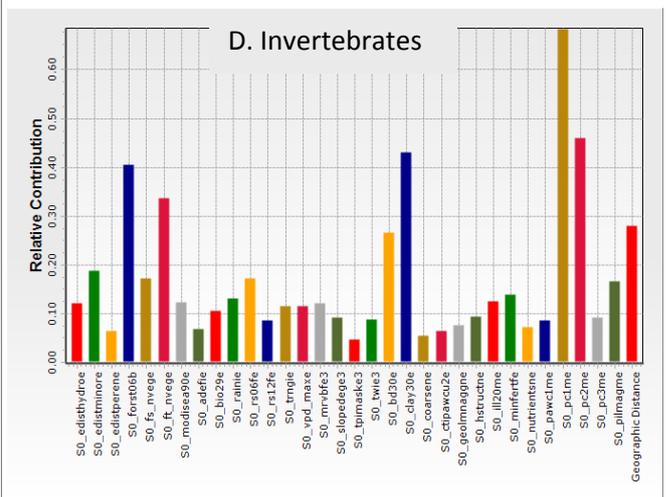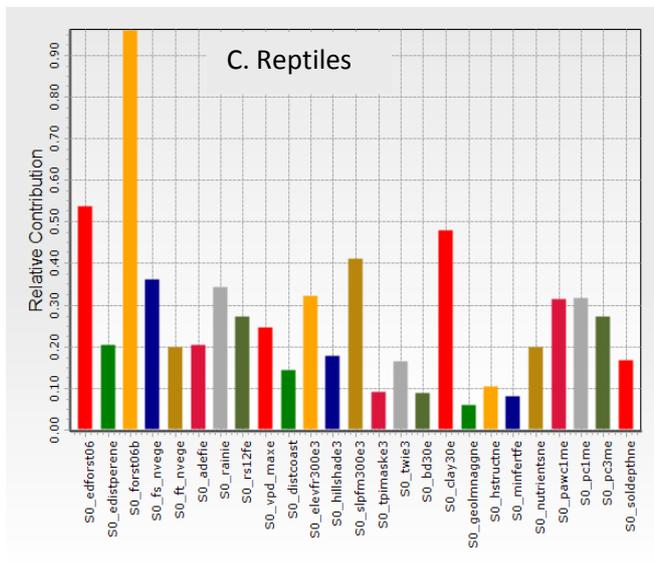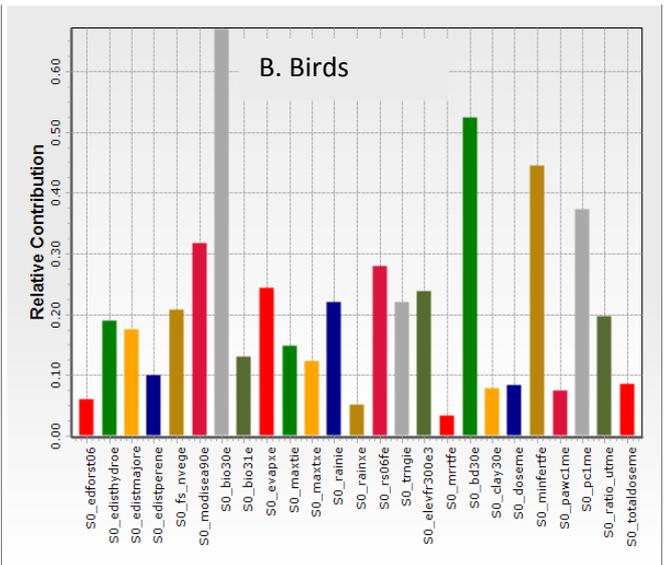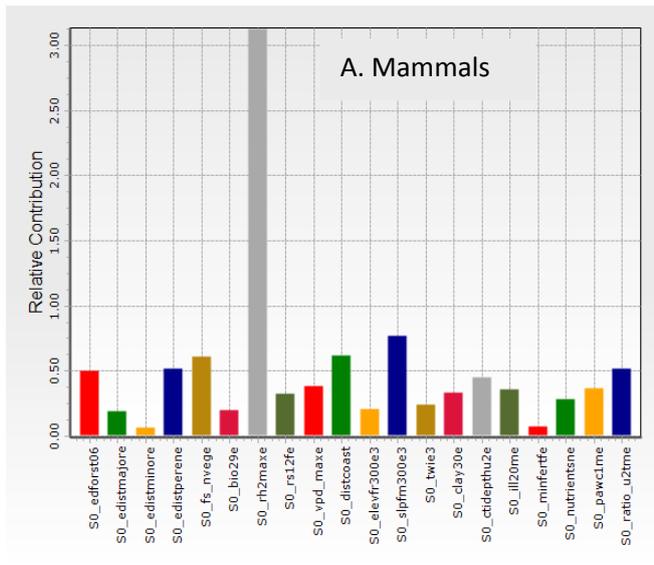
Figure 44. Histogram showing the relative contribution (sum of all spline coefficient values) of each predictor in the fitted model for the five biological groups. Predictor variables are described in Appendix B .
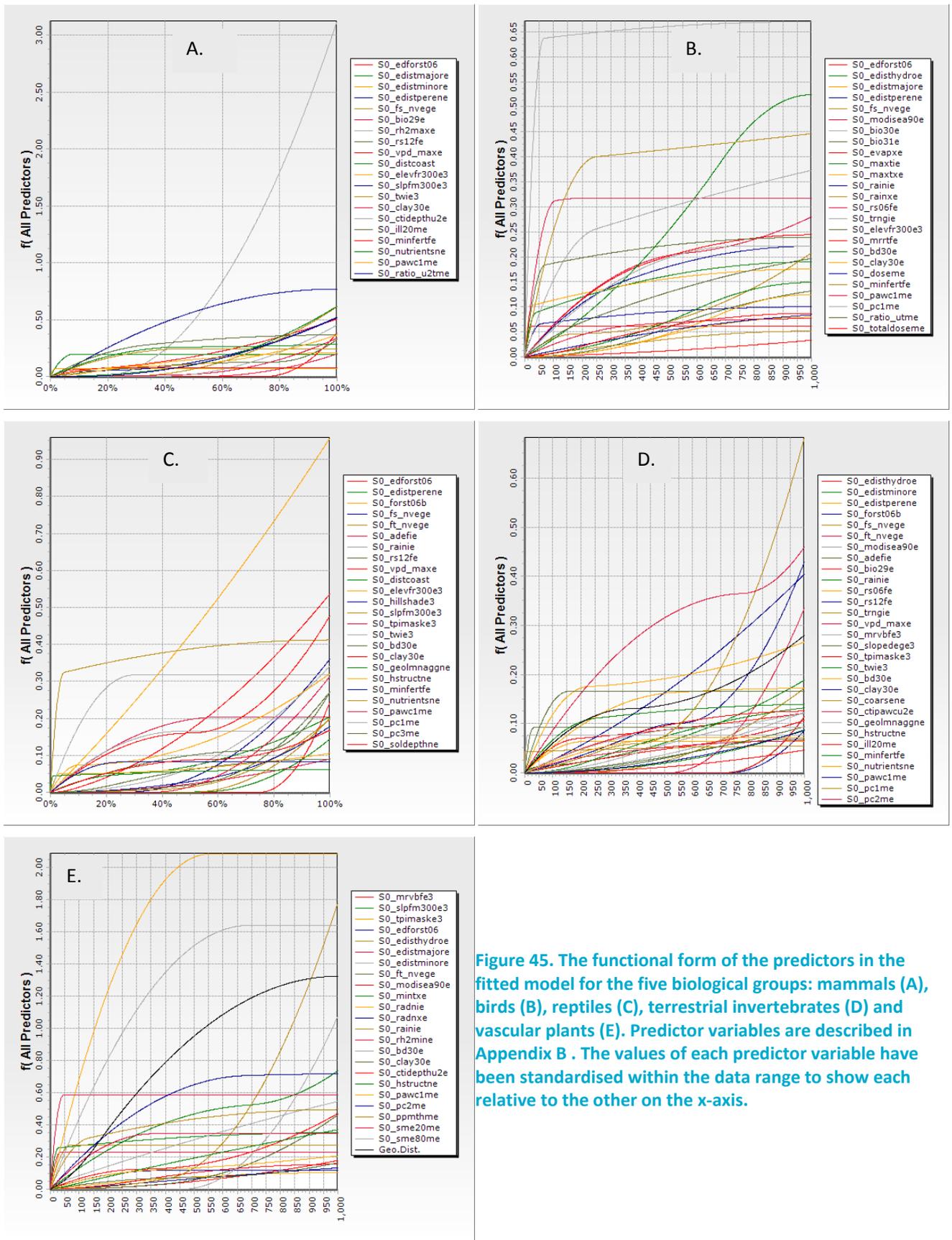
Figure 45. The functional form of the predictors in the fitted model for the five biological groups: mammals (A), birds (B), reptiles (C), terrestrial invertebrates (D) and vascular plants (E). Predictor variables are described in Appendix B . The values of each predictor variable have been standardised within the data range to show each relative to the other on the x-axis.

Appendix D Richness models: Random Forests, GAM

6.5 Random Forests Models

Exploratory Random Forest models identified the set of candidate predictor variables that have the most explanatory power (Table 21, Figure 46 to Figure 50). To interpret the random forest plot, briefly, the importance measures show how much mean squared error (MSE) node impurity increases when a variable is randomly permuted. If we randomly permute a variable that does not result in a gain in prediction, then predictions will not change greatly and we see only small changes in MSE and node impurity. On the other hand, the important variables will change the predictions by a large amount when randomly permuted, and so we see bigger changes in MSE and node impurity. This indicates variable importance. The Random Forest models also provide a baseline expectation for the percent variance explained by the predictors given the data.

Table 21: Summary statistics for the Random Forest models.

BIOLOGICAL GROUP	MODEL TYPE	NUMBER OF TREES	VARIABLES TRIED EACH SPLIT	MEAN OF SQUARED RSIDUALS	% VARIANCE EXPLAINED	BEST PREDICTOR[1]	NUMBER OF PREDICTORS SELECTED
Mammals	Regression	500	17	1.529	24.6	VPD_MAX	11
Birds	Regression	500	32	95.03	21.88	EDISTPEREN	11
Reptiles	Regression	500	16	21.32	32.67	RAINX	12
Invertebrates	Regression	500	31	59.92	27.85	EDISTHYDRO	10
Vascular plants	Regression	500	21	122.4	31.87	KAO20	12

1. VPD_MAX - Maximum monthly vapour pressure deficit (KPa); EDISTPEREN - Euclidean distance to perennial water bodies and drainage; RAINX - maximum monthly rainfall (mm); EDISTHYDRO - Euclidean distance to water points; KAO20 - % abundance modelled kaolinite clay minerals in surface soil (0-20cm). Further details are given in Appendix B .
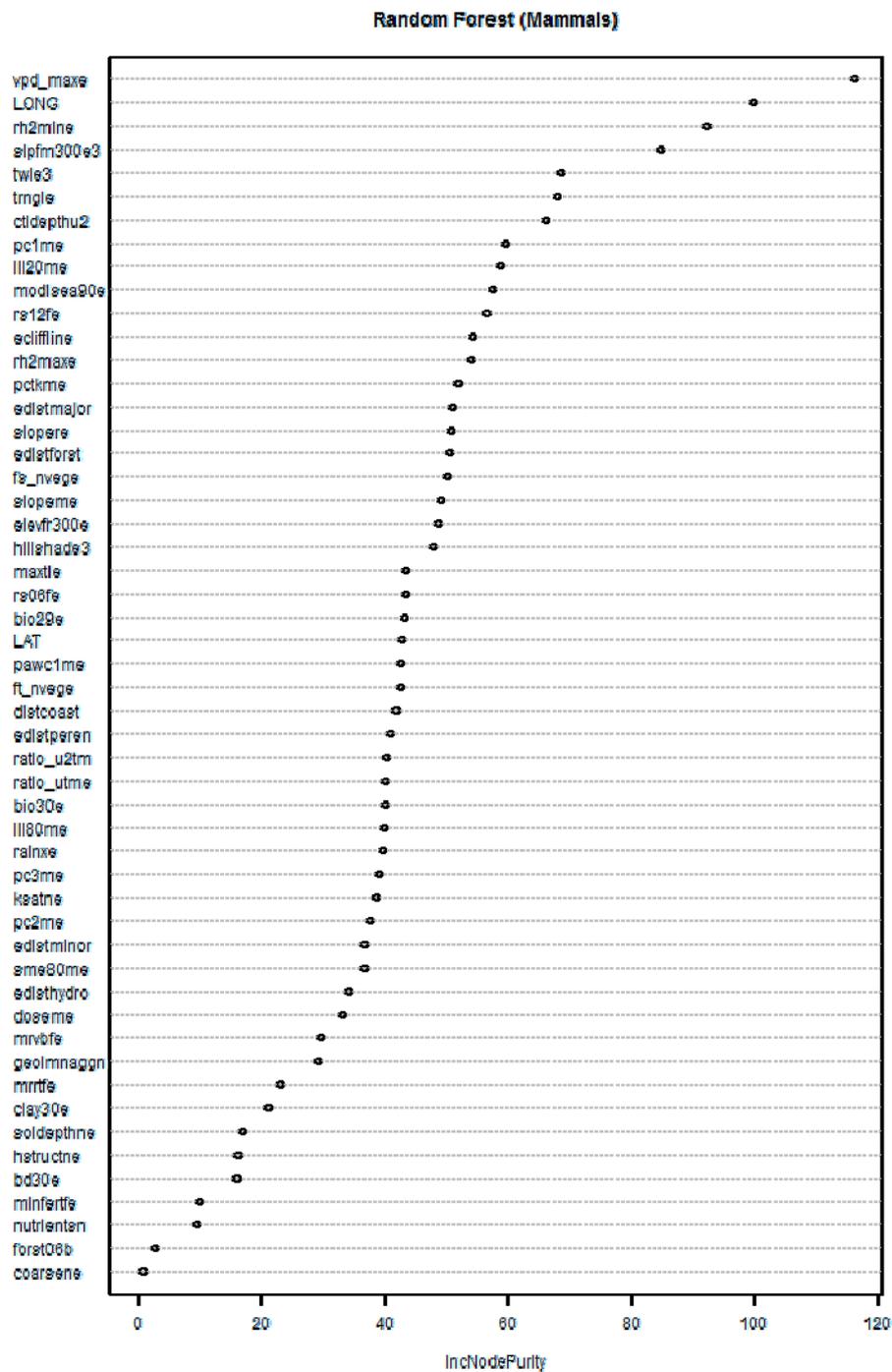
Figure 46. Predictor variable importance for mammal richness using Random forests. Selected variables are: VPD_MAXE, LONG, RH2MINE, SLPFM300E3, TWIE3, TRNGIE, CTIDEPTHU2, PC1ME, ILL20ME, MODISEA90E, RS12FE.
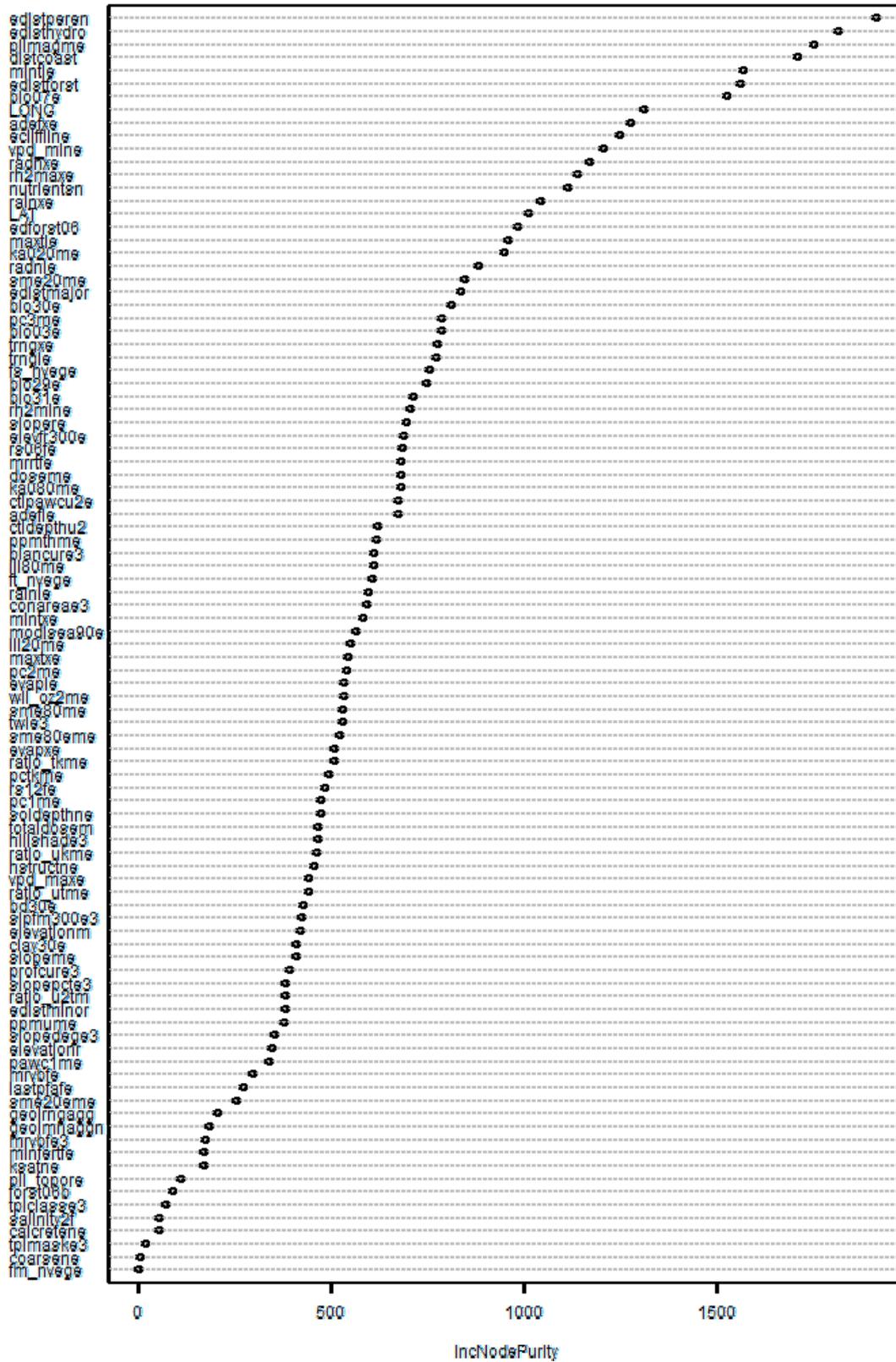
Figure 47. Predictor variable importance for bird richness using Random forests. Selected variables are: EDISTPEREN, EDISTHYDRO, PILMAGME, DISTCOAST, MINTIE, EDISTFORST, BIO07E, LONG, ADEFXE, ECLIFFLINE, VPD_MINE.
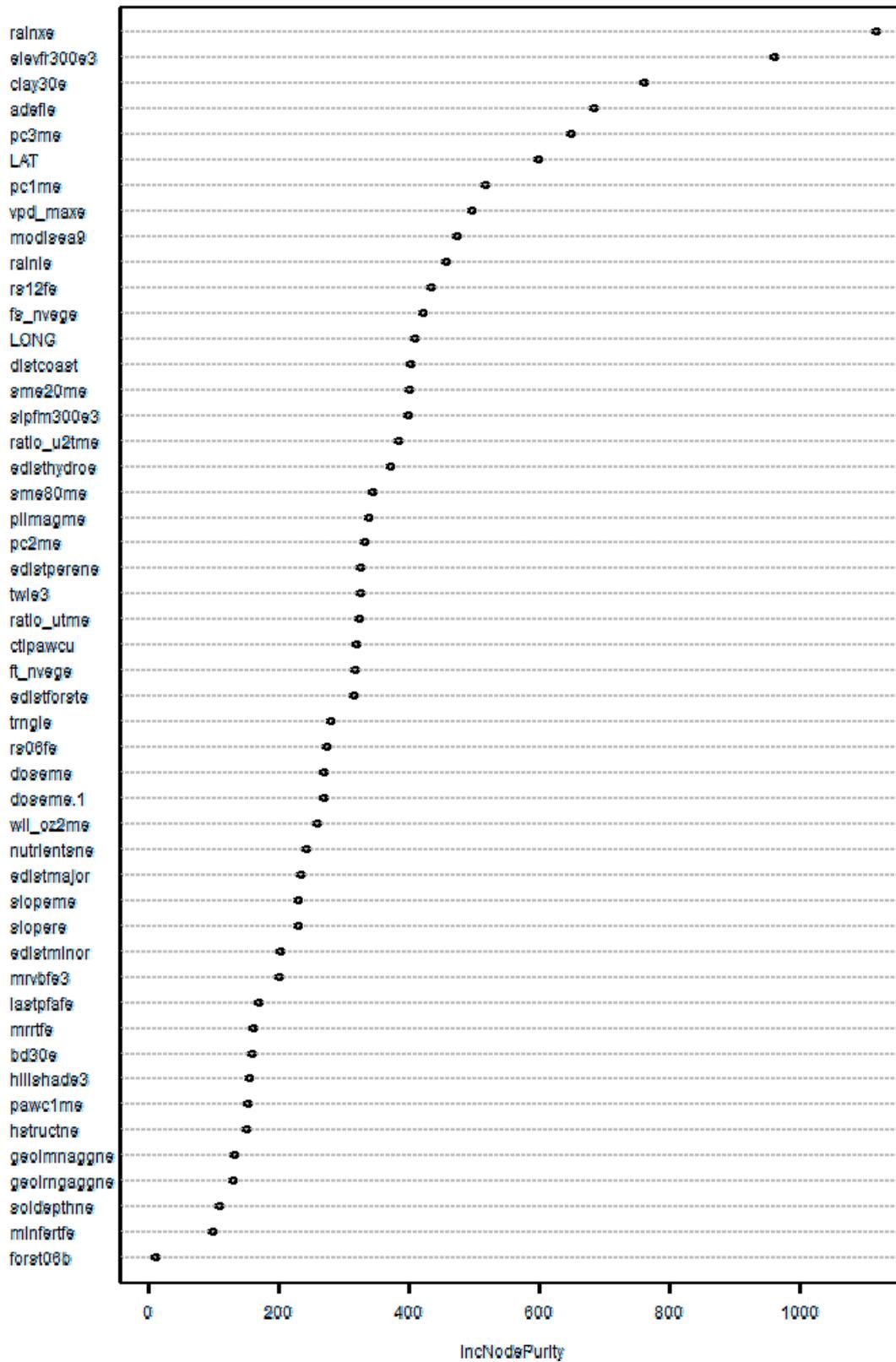
Figure 48. Predictor variable importance for reptiles richness using Random forests. Selected variables are: RAINXE, ELEVFR300E3, CLAY30E, ADEFIE, PC3ME, LAT, PC1ME, VPD_MAXE, MODISEA9, RAINIE, RS12FE, FS_NVEGE.
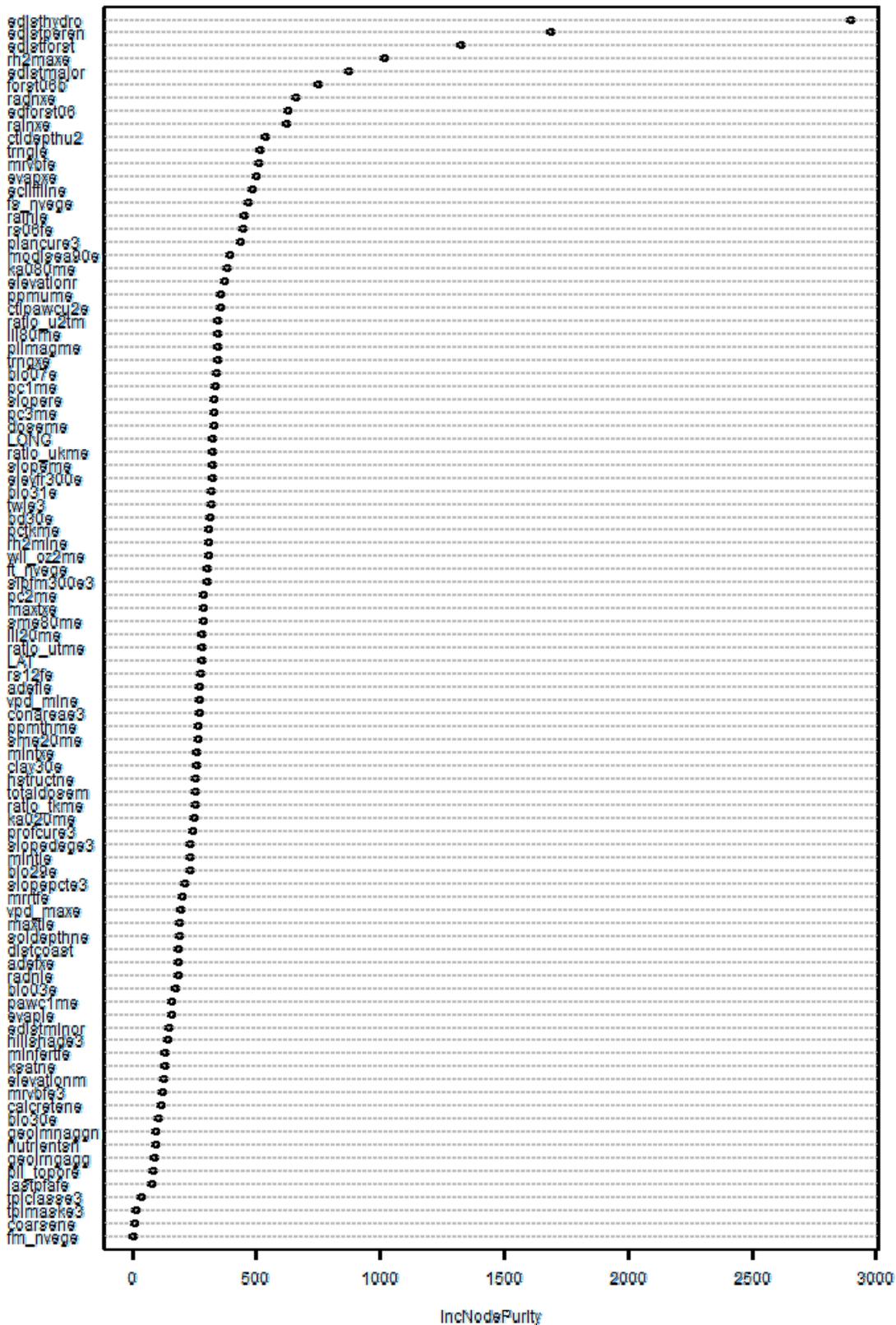
Figure 49. Predictor variable importance for invertebrate richness using Random forests. Selected variables are: EDISTHYDRO, EDISTPEREN, EDISTFORST, RH2MAXE, EDISTMAJOR, FORST06B, RADNXE, EDFORST06, RAINXE, CTIDEPTHU2.
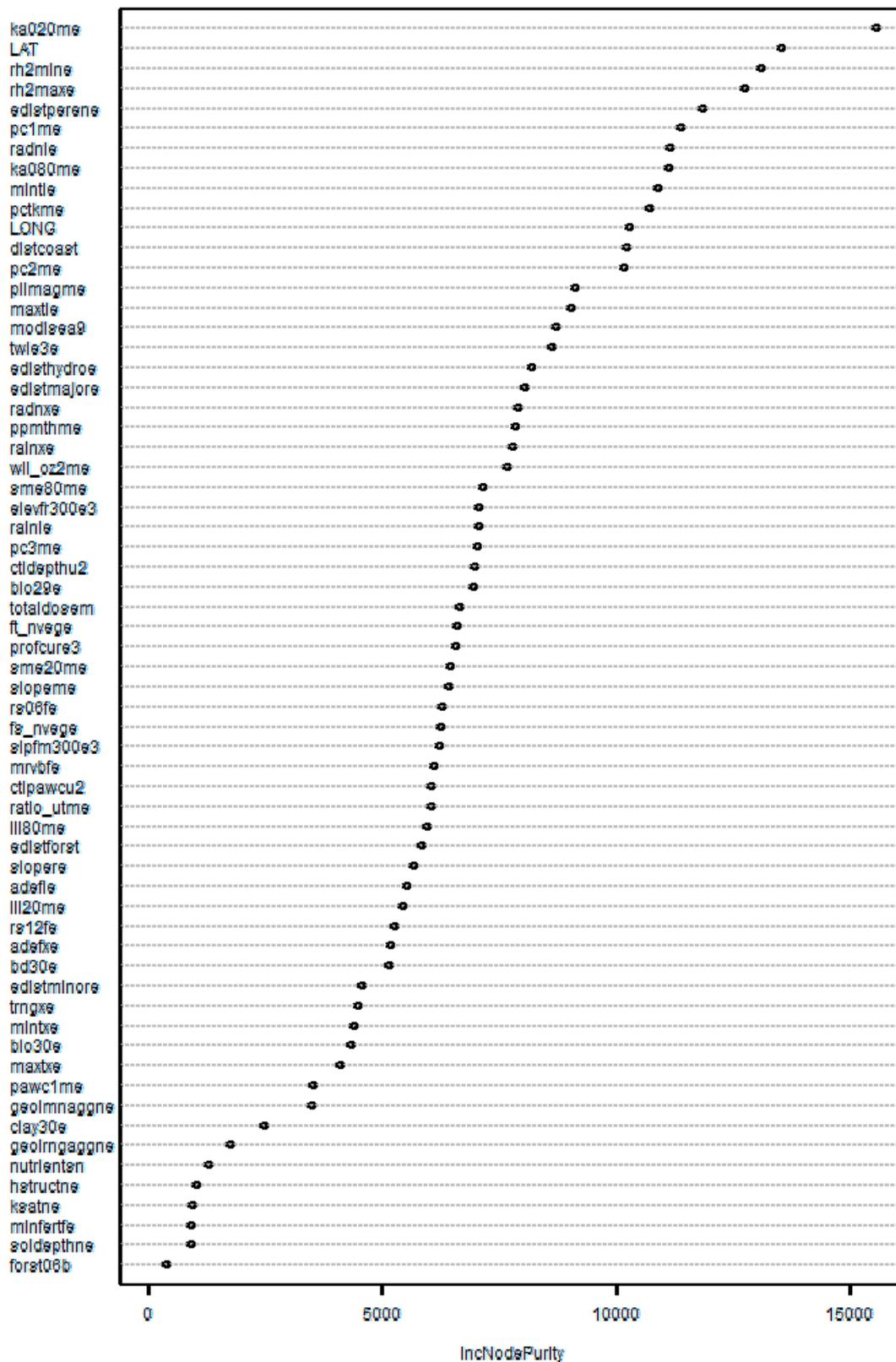
Figure 50. Predictor variable importance for vascular plant richness using Random forests. Selected variables are: KA020ME, LAT, RH2MINE, RH2MAXE, EDISTPERENE, PC1ME, RADNIE, KA080ME, MINTIE, PCTKME, LONG, DISTCOAST.

6.6 Generalised Additive Models

6.6.1 RICHNESS MODEL FOR MAMMALS

The model does not perform well when interaction terms are included, so the model used for prediction is:

```
Family: gaussian
Link function: identity

Formula:
SR ~ s(vpd_maxe, k = 16) + s(rh2mine, k = 12) + s(slpfm300e3,
    k = 8) + s(trngie, k = 12) + s(ctidepthu2, k = 12) + s(pc1me,
    k = 8) + s(ill20me, k = 10) + s(slopere, k = 8)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.7868     0.0356    50.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Approximate significance of smooth terms:
              edf Ref.df     F p-value
s(vpd_maxe)  10.99  12.74 2.30 0.00559 **
s(rh2mine)    9.24  10.20 4.28 6.1e-06 ***
s(slpfm300e3) 5.16   6.03 8.72 2.4e-09 ***
s(trngie)     9.81  10.56 5.17 1.1e-07 ***
s(ctidepthu2) 8.40   9.74 2.36 0.01004 *
s(pc1me)      6.30   6.82 3.54 0.00105 **
s(ill20me)    7.68   8.56 3.31 0.00071 ***
s(slopere)    3.99   4.55 2.20 0.05910 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.223   Deviance explained = 26.2%
GCV score = 1.6594  Scale est. = 1.5759    n = 1243


Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                 k'      edf k-index p-value
s(vpd_maxe)   15.000 10.989   0.970    0.14
s(rh2mine)    11.000  9.243   0.914    0.00
s(slpfm300e3)  7.000  5.159   0.944    0.03
s(trngie)     11.000  9.814   0.955    0.06
s(ctidepthu2) 11.000  8.398   0.895    0.00
s(pc1me)       7.000  6.300   0.929    0.01
s(ill20me)     9.000  7.685   0.970    0.14
s(slopere)     7.000  3.993   0.972    0.14
```

GCV (generalised cross validation) is based on minimizing some approximation to the expected prediction error of the model. When a predictor is dropped, GCV is expected to increase requiring a balance be found between the simplicity of the model and the errors. The predictors in the resulting model are chosen by optimising GCV score and hence the GCV method of smoothing parameters in the generalised additive model with multiple penalties. To achieve this (optimising GCV score) the 'magic' function adopts Newton's method and steepest descent iteratively to adjust the smoothing parameters for each penalty term.
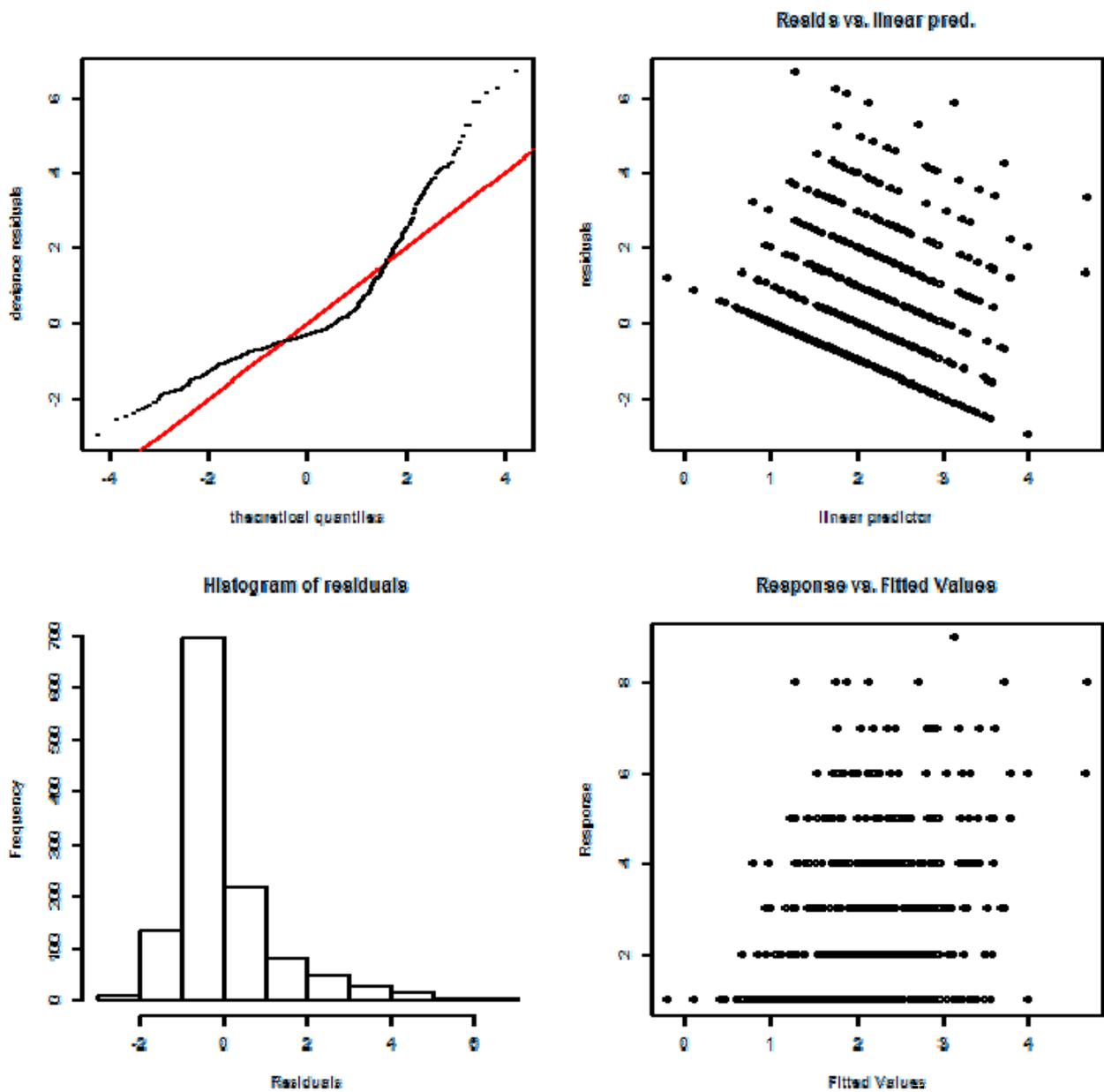
Figure 51. A graphical evaluation of the statistical assumptions of the model fitted to the richness dataset for mammals. The upper left normal QQ plot departs from relatively a straight line, suggesting that the distributional assumption is inconsistent with the data. The upper right plot suggests that variance is approximately constant as the mean increases. The histogram of residuals at lower left shows the departure from normality (left skewed). The lower right plot of response against fitted values shows a positive linear relationship but the model under predicts richness.
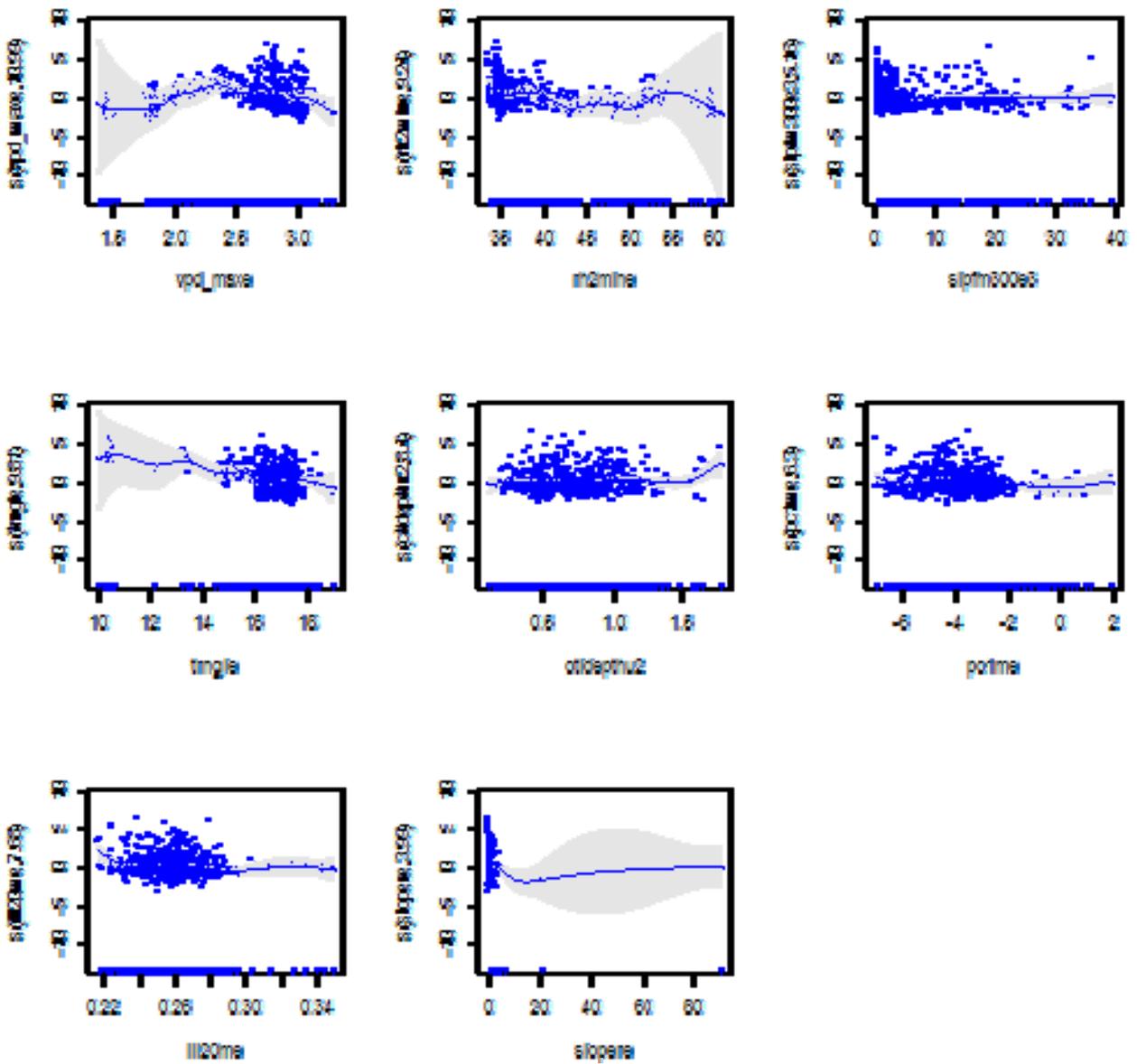
Figure 52. The shape of the predictor fitted function in the richness model for mammals.

6.6.2 RICHNESS MODEL FOR BIRDS

Interactions were found to improve the model performance and were included:

```
Family: gaussian
Link function: identity

Formula:
SR ~ s(distcoast, edistforst, k = 16) + s(edistforst, mintie,
    k = 16) + s(pilmagme, k = 2) + edistforst + s(edisthydro,
    k = 6) + s(bio07e, ecliffline, k = 16) + s(distcoast, adefxe,
    k = 6)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   13.866      0.378    36.7   <2e-16 ***
edistforst     1.447      0.125    11.6   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                        edf Ref.df     F p-value
s(distcoast,edistforst) 13.89  14.73 7.12 3.4e-14 ***
s(edistforst,mintie)     7.11   9.03 1.68  0.0906 .
s(pilmagme)              1.76   1.94 1.76  0.1725
s(edisthydro)            3.75   4.33 1.77  0.1274
s(bio07e,ecliffline)    13.63  14.52 3.76 3.4e-06 ***
s(distcoast,adefxe)      3.98   4.00 3.96  0.0036 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.342   Deviance explained = 39.4%
GCV score =  87.39  Scale est. = 80.231    n = 551


Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                        k'      edf k-index p-value
s(distcoast,edistforst) 15.000 13.892   0.951    0.11
s(edistforst,mintie)    14.000  7.112   1.004    0.53
s(pilmagme)              2.000  1.760   1.018    0.63
s(edisthydro)            5.000  3.746   0.971    0.24
s(bio07e,ecliffline)    15.000 13.631   1.006    0.54
s(distcoast,adefxe)      4.000  3.984   1.030    0.78
```
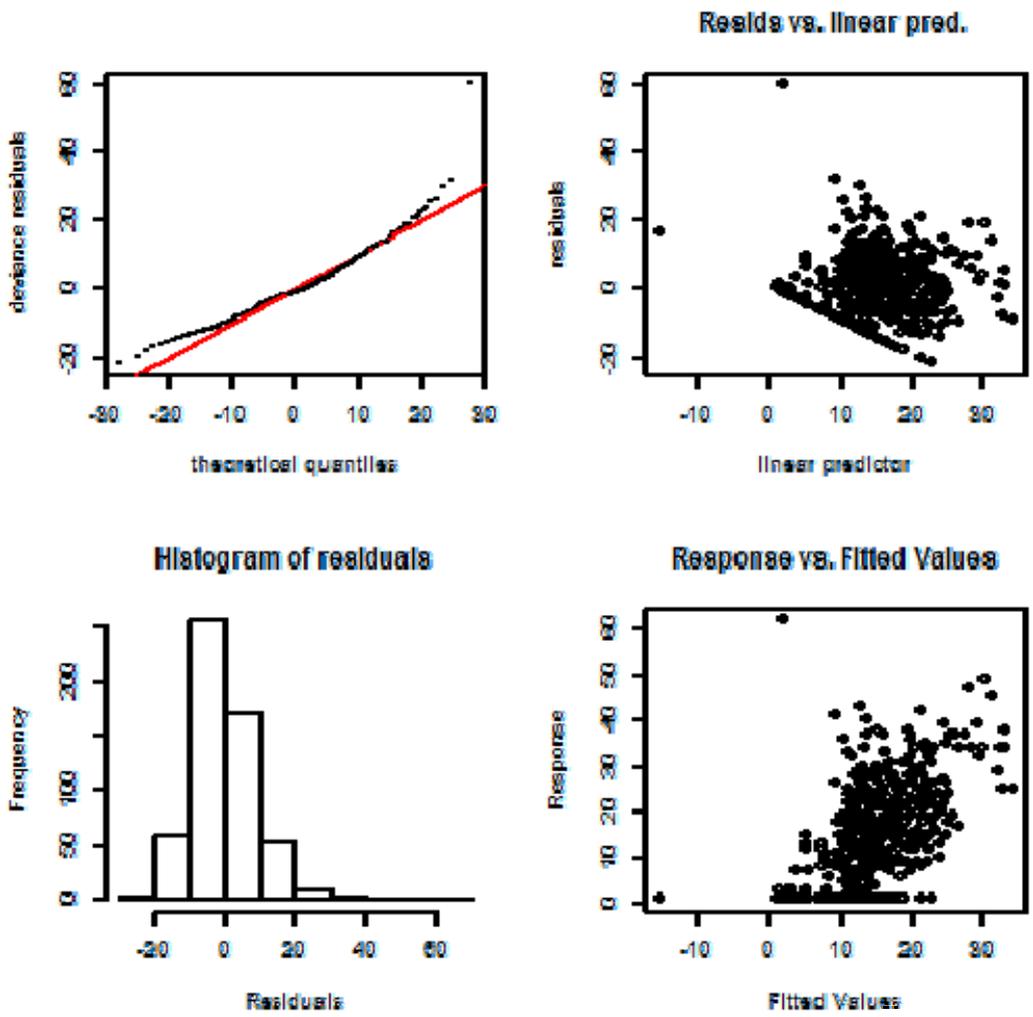
Figure 53. A graphical evaluation of the statistical assumptions of the model fitted to the richness dataset for birds. The upper left normal QQ plot is relatively close to a straight line, suggesting that the distributional assumption is reasonable, however it is slightly skewed with a longer tail on the right. The upper right plot suggests that variance is approximately constant as the mean increases. The histogram of residuals at lower left appears approximately consistent with normality (although slightly skewed and some asymmetry). The lower right plot of response against fitted values shows a positive linear relationship. The response data are integers, this is why we see a straight line at the bottom of the residual plot, it corresponds to the value 1.
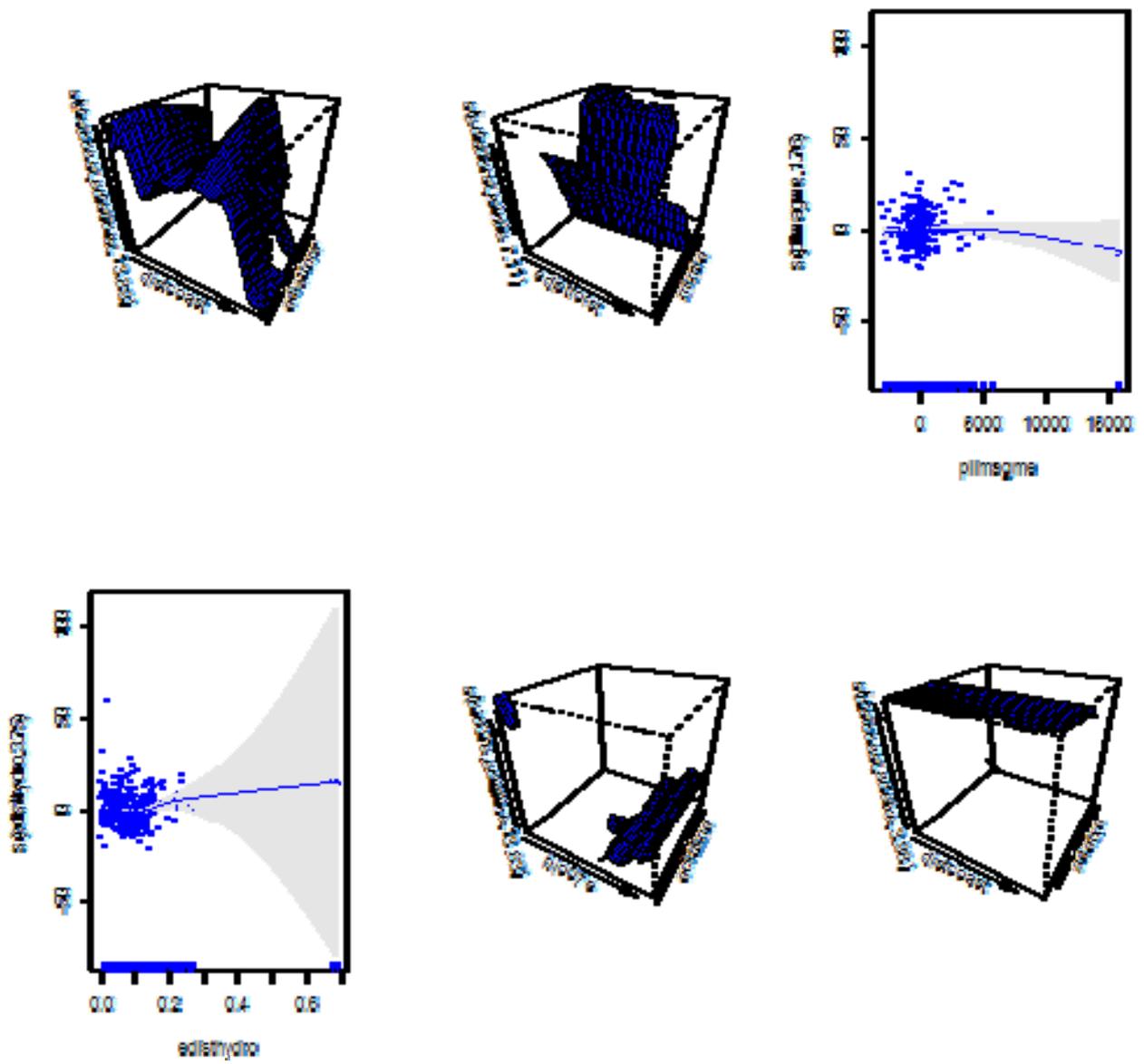
Figure 54. The shape of the predictor fitted functions and interactions in the richness model for birds.

6.6.3 RICHNESS MODEL FOR REPTILES

Interactions were found to improve the model performance and were included:

```
Family: gaussian
Link function: identity

Formula:
SR ~ s(elevfr300e3, rainxe, clay30e, k = 16) + s(adefie, k = 4) +
    s(pc3me, k = 6) + vpd_maxe + s(fs_nvege, k = 16) + s(rainie,
    rainxe, k = 8) + rs12fe

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   33.383      8.585    3.89  0.00011 ***
vpd_maxe       2.562      1.396    1.83  0.06711 .
rs12fe        -1.221      0.394   -3.10  0.00204 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                            edf Ref.df    F p-value
s(elevfr300e3,rainxe,clay30e) 10.31  11.28 5.17 7.4e-08 ***
s(adefie)                      2.78   2.96 2.72 0.04466 *
s(pc3me)                       2.38   3.02 2.66 0.04736 *
s(fs_nvege)                   12.45  14.01 2.68 0.00083 ***
s(rainie,rainxe)               6.90   6.99 6.81 9.4e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.336   Deviance explained =   38%
GCV score = 22.587  Scale est. = 21.055    n = 558


Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                            k'     edf k-index p-value
s(elevfr300e3,rainxe,clay30e) 14.000 10.315   1.047    0.90
s(adefie)                      3.000  2.782   0.994    0.42
s(pc3me)                       5.000  2.382   1.025    0.67
s(fs_nvege)                   15.000 12.454   1.002    0.54
s(rainie,rainxe)               7.000  6.899   1.030    0.76
```
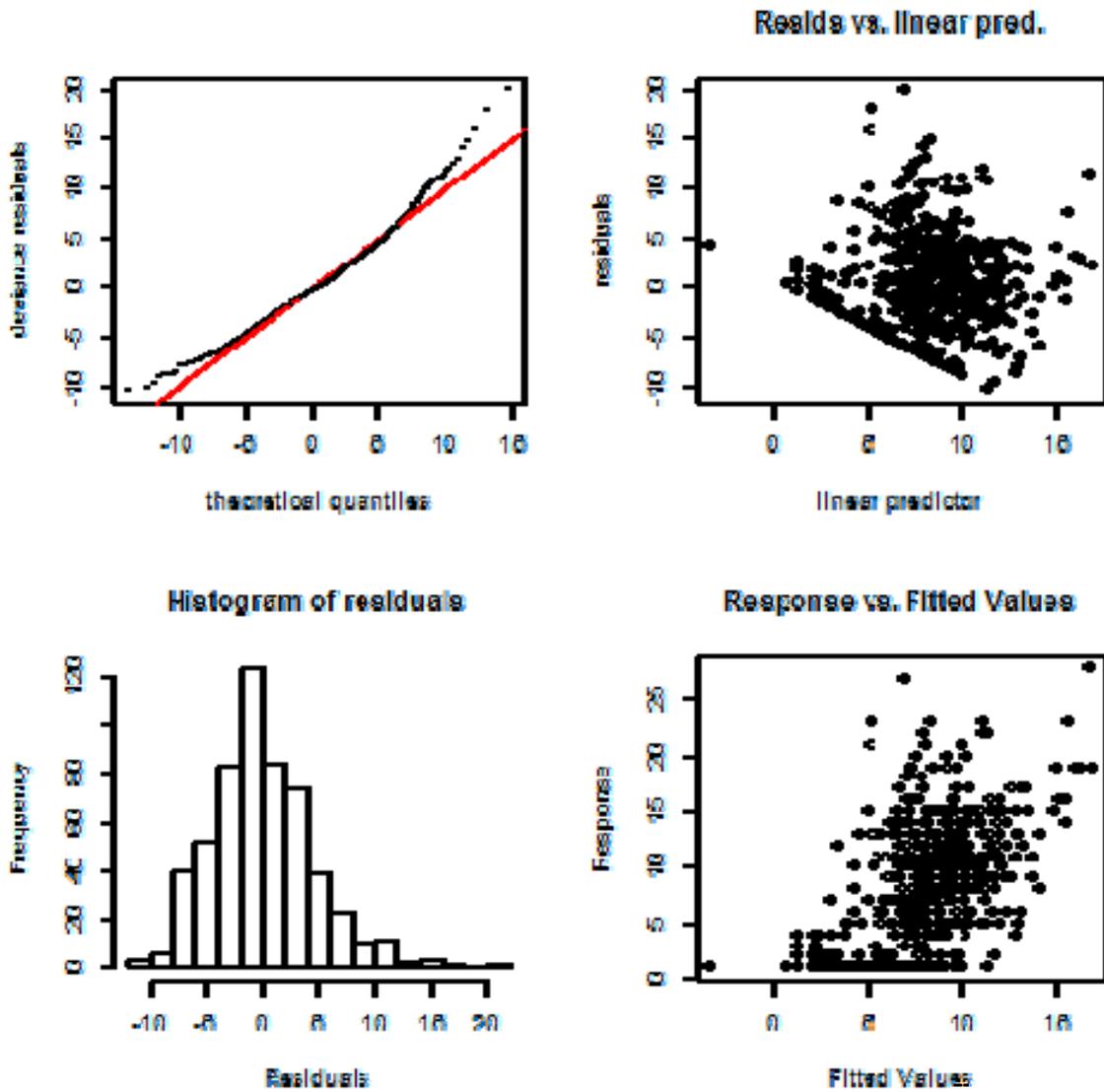
Figure 55. A graphical evaluation of the statistical assumptions of the model fitted to the richness dataset for reptiles. The upper left normal QQ plot is relatively close to a straight line, suggesting that the distributional assumption is reasonable, however it is slightly skewed with a longer tail on the right. The upper right plot suggests that variance is approximately constant as the mean increases. The histogram of residuals at lower left appears approximately consistent with normality (although slightly skewed with some asymmetry). The lower right plot of response against fitted values shows a positive linear relationship.
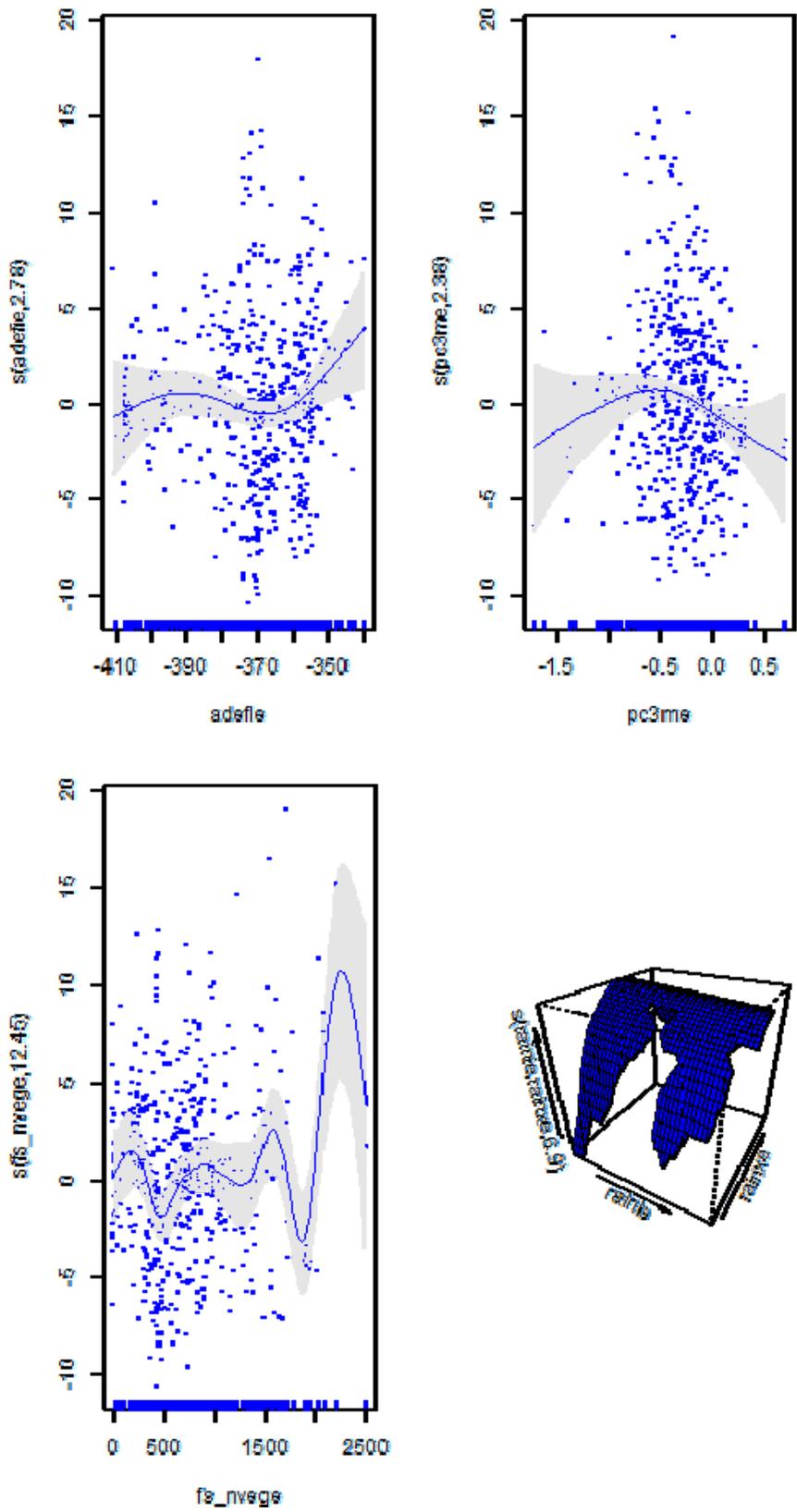
Figure 56. The shape of the predictor fitted functions and interactions in the richness model for reptiles.

6.6.4 RICHNESS MODEL FOR INVERTEBRATES

Interactions were found to improve the model performance and were included:

```
Family: gaussian
Link function: identity

Formula:
SR ~ s(edisthydro, edistperen, k = 16) + forst06b + ctidepthu2

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.83e+01    1.10e+00   16.58  < 2e-16 ***
forst06b    8.58e-04    1.81e-04    4.73  3.1e-06 ***
ctidepthu2  4.14e+00    1.34e+00    3.08   0.0022 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                          edf Ref.df    F p-value
s(edisthydro,edistperen) 8.73   11.1 8.47 9.5e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.271   Deviance explained = 29.1%
GCV score = 62.492  Scale est. = 60.673    n = 403


Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                          k'    edf k-index p-value
s(edisthydro,edistperen) 15.00  8.73    0.95    0.13
```
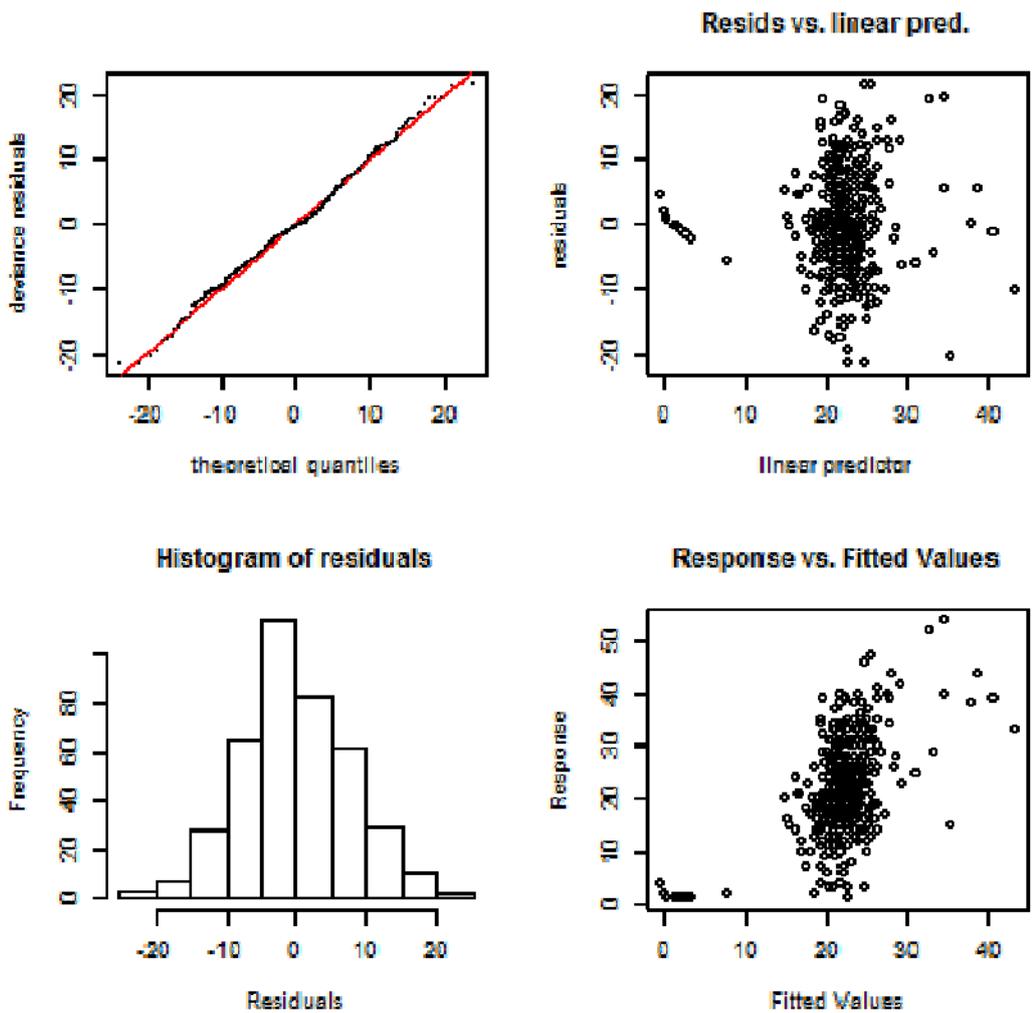
Figure 57. A graphical evaluation of the statistical assumptions of the model fitted to the richness dataset for invertebrates. The upper left normal QQ plot is relatively close to a straight line, suggesting that the distributional assumption is reasonable, however it is slightly skewed with a longer tail on the right. The upper right plot suggests that variance is approximately constant as the mean increases. The histogram of residuals at lower left appears consistent with normality. The lower right plot of response against fitted values shows a positive linear relationship.
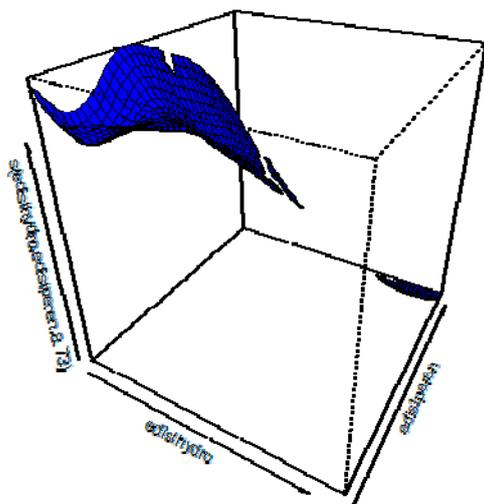
Figure 58. The shape of the predictor fitted functions and interactions in the richness model for invertebrates.

6.6.5 RICHNESS MODEL FOR VASCULAR PLANTS

Interactions were found to improve the model performance and were included:

```
Family: gaussian
Link function: identity

Formula:
SR ~ s(pctkme, k = 4) + s(rh2mine, k = 12) + s(ka020me, k = 12) +
    s(rh2maxe, k = 12) + s(ka020me, edistperene, pc1me, k = 16) +
    s(ka080me, k = 6) + s(radnie, k = 12) + s(pc2me, k = 16)

Parametric coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   22.418      0.241    93.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
                              edf Ref.df     F p-value
s(pctkme)                    2.49   2.82  8.59 2.4e-05 ***
s(rh2mine)                  10.51  10.94  4.41 1.5e-06 ***
s(ka020me)                   9.01   9.86  3.44 0.00019 ***
s(rh2maxe)                  10.60  10.96  6.04 9.1e-10 ***
s(ka020me,edistperene,pc1me) 9.56  10.40  6.16 1.5e-09 ***
s(ka080me)                   2.76   3.45  2.54 0.04707 *
s(radnie)                   10.22  10.85  9.76 < 2e-16 ***
s(pc2me)                     9.49  11.47  5.23 2.1e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.197   Deviance explained = 21.8%
GCV score = 148.21  Scale est. = 144.3    n = 2492


Basis dimension (k) checking results. Low p-value (k-index<1) may
indicate that k is too low, especially if edf is close to k'.

                                k'    edf k-index p-value
s(pctkme)                    3.000  2.492   0.947    0.00
s(rh2mine)                  11.000 10.509   0.939    0.00
s(ka020me)                  11.000  9.012   1.002    0.54
s(rh2maxe)                  11.000 10.601   0.957    0.02
s(ka020me,edistperene,pc1me) 14.000  9.565   0.922    0.00
s(ka080me)                   5.000  2.764   0.954    0.01
s(radnie)                   11.000 10.220   0.919    0.00
s(pc2me)                    15.000  9.493   0.988    0.28
```
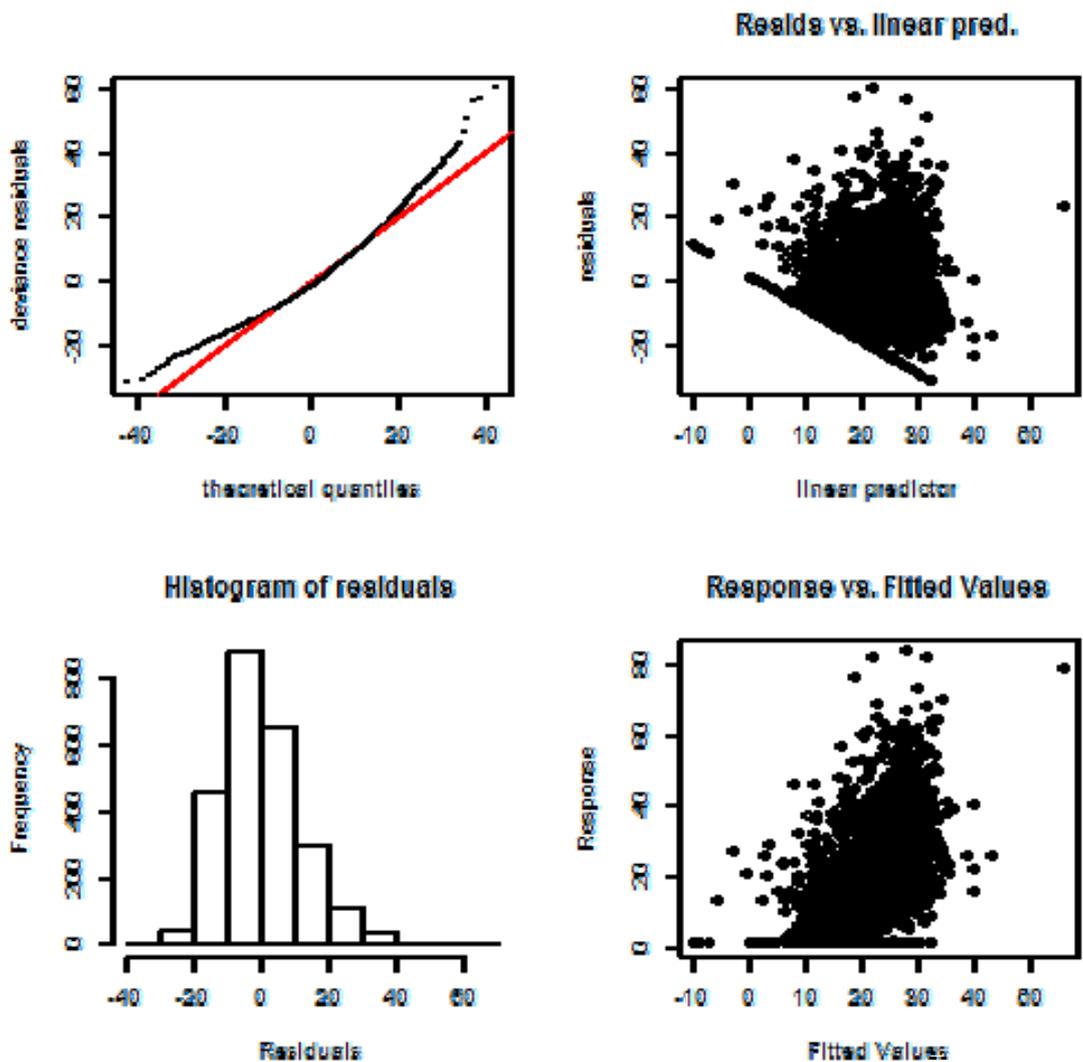
Figure 59. A graphical evaluation of the statistical assumptions of the model fitted to the richness dataset for vascular plants. The upper left normal QQ plot is relatively close to a straight line, suggesting that the distributional assumption is reasonable, however it is slightly skewed with a longer tail on the right. The upper right plot suggests that variance is approximately constant as the mean increases. The histogram of residuals at lower left appears approximately consistent with normality (although slightly skewed with some asymmetry). The lower right plot of response against fitted values shows a positive linear relationship. The response data are integers, this is why we see a straight line at the bottom of the residual plot; it corresponds to the value 1.
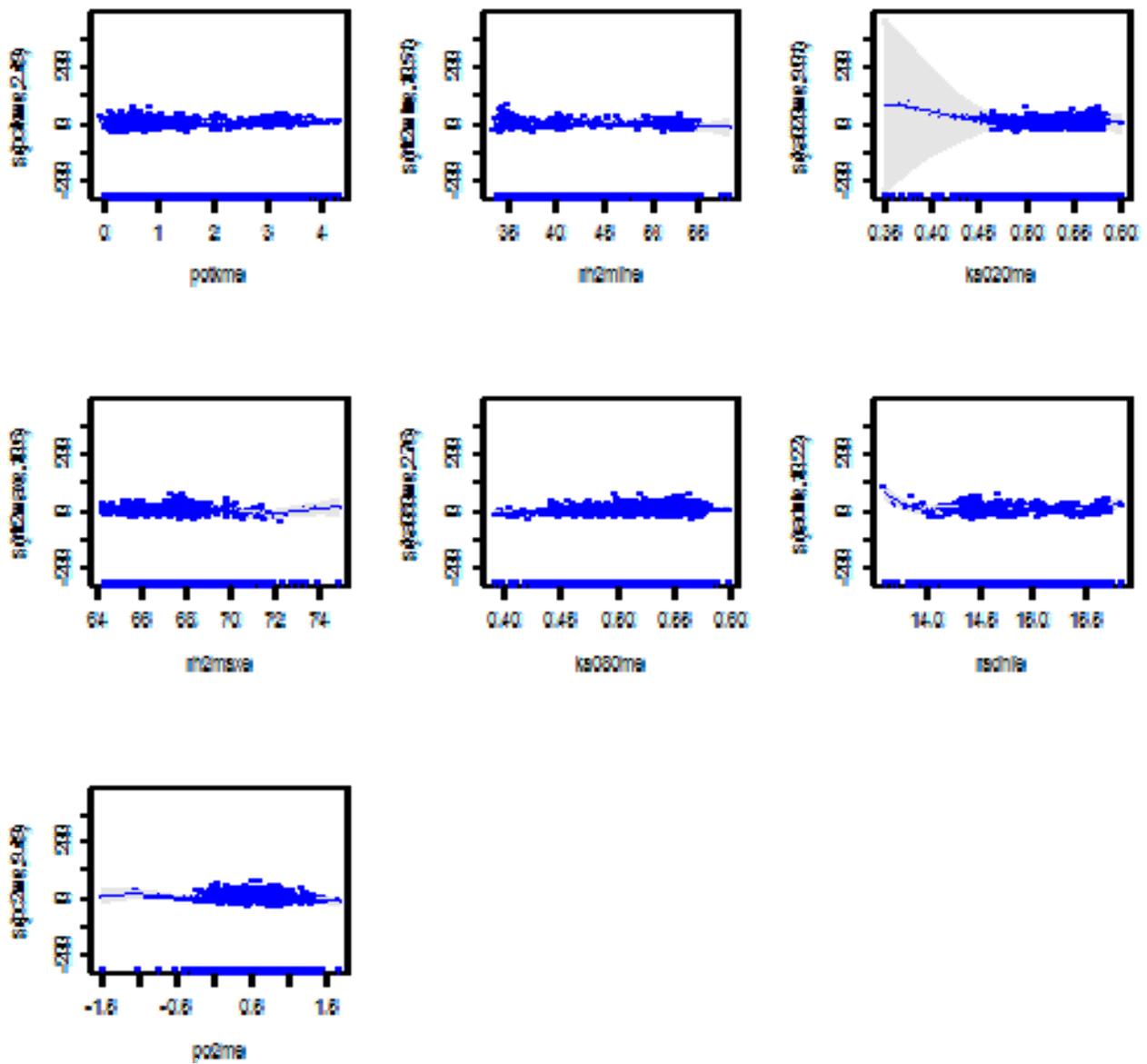
Figure 60. The shape of the predictor fitted functions (and interactions – not shown) in the richness model for vascular plants.

Appendix E Analysis of vegetation cover change

In this appendix we present a significant analysis of vegetation cover change based on an assessment of satellite-based data, provide examples of its utility and outline the requirements for refining these data to inform biodiversity habitat condition mapping in the future.

This assessment used satellite-derived bare ground values over a 10-year period to derive mean bare ground and cover change indices. The fine scaled analysis (approximately 30m pixels) and has significant promise for future refinement and validation of condition. In the absence of contextual management data to use in calibrating the change analysis with ecosystem condition, these outputs were not ready to apply in the subsequent biodiversity significance assessment. We therefore undertook a rapid assessment of habitat condition at a broader resolution using the best-available, accessible datasets for land use, tenure and infrastructure as indicators of the potential for human modification of ecosystem function (Section 4.3). These outputs will likely also prove useful in validating the remotely-sensed assessment of condition.

6.7 Ground cover analysis from satellite data

To complete the temporal analysis of ground cover in the Pilbara region, 10 years of ground cover indices were downloaded from (AusCover 2013). A combination of Linux shell script and code written in R (2.15.0) (R Development Core Team 2009) was used to search the AusCover website for all of the dry season dates for each relevant scene (2000 – 2011) (Table 22, Figure 61) and down load these to a local server.

The AusCover fractional cover imagery are pre-processed, masked ground cover percentage grids derived from Landsat TM and ETM+L2T acquired from the USGS (2013) who completed calibration and geo-registration. Dry season dates where acquired to increase the chance of distinguishing between evergreen vegetation and ground cover. Images are corrected using the methods outlined in Danaher *et al.* (2010) and were tested for accuracy against a nationwide field sampling network following published survey methods (Muir *et al.* 2011). Confounding temporal errors such as cloud shadow, topographic shadow and water were masked from the images. The fractional cover grid is measured in percentages of cover, here we used band 1 (percentage of bare ground, rock or disturbed ground).

Table 22. Path and row combinations of fractional cover scenes used in the analysis.

PATH	ROW	DATUM	ZONE	EPSG	DATES
111	074	WGS84	51s	32751	2000-2010
111	075	WGS84	51s	32751	2000-2010
111	076	WGS84	51s	32751	2000-2010
112	074	WGS84	50s	32750	2000-2010
112	075	WGS84	50s	32750	2000-2010
112	076	WGS84	50s	32750	2000-2010
113	074	WGS84	50s	32750	2000-2010
113	075	WGS84	50s	32750	2000-2010

113	076	WGS84	50s	32750	2000-2010
114	074	WGS84	50s	32750	2000-2010
114	075	WGS84	50s	32750	2000-2010
114	076	WGS84	50s	32750	2000-2010
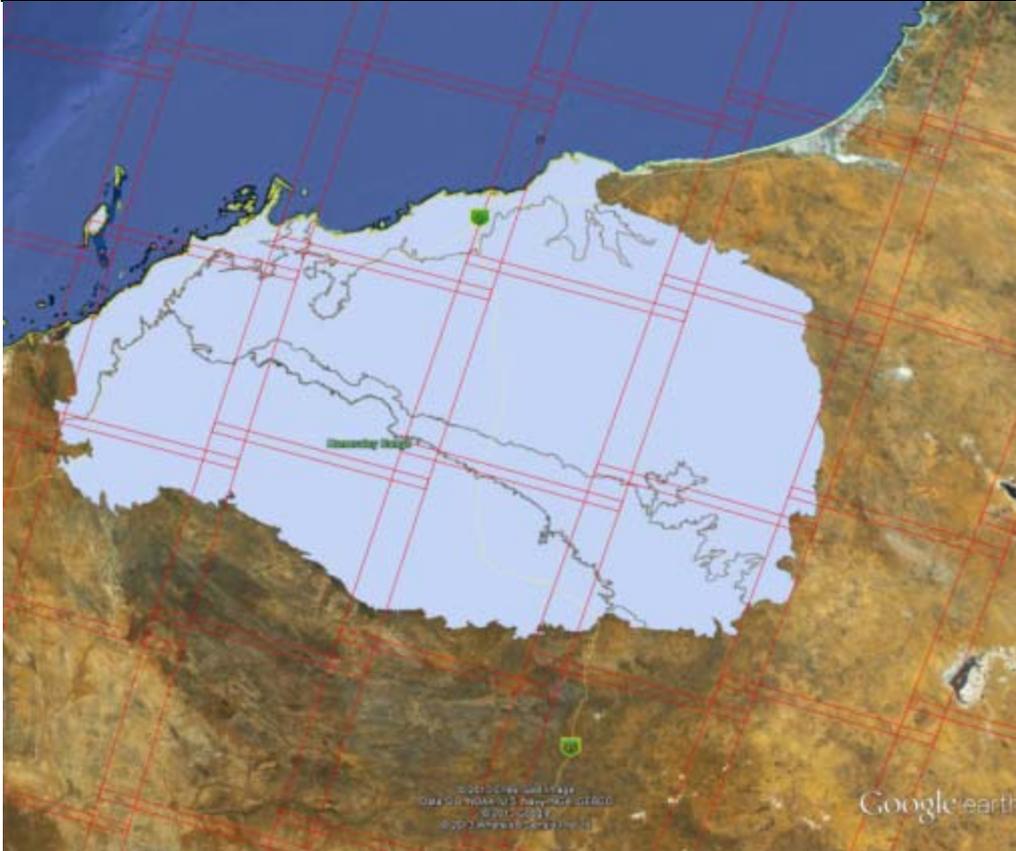
Figure 61. Area of interest for the fractional cover analysis (blue polygon) illustrating the coverage of Landsat scenes (red outline).

Pre-processed fractional cover grids were provided as netcdf files and were converted to ascii grids for analysis using the 'ncdf' package (Pierce 2013) in R (2.15.0) (R Development Core Team 2009). A total of 120 grids were converted comprising 12 path row combinations and 10 years of temporal data. Due to the large size of each grid and the large number of scenes processed, a script was written to run each scene in parallel on the James Cook University (JCU) High Performance Computing portal (HPC) in Townsville.

6.7.1 PROCESSING THE SATELLITE DATA

Once converted to ascii format, the 'SDM Tools' package (VanDerWal *et al.* 2012) in R was used to convert the grids into a data frame including the location of each grid cell (easting and northing) and the bare ground percentage. To analyse the change in percentage bare ground and to derive a mean bare ground index, a matrix of values for each pixel in each year was required. The pixel size was consistent across time but the extent changed slightly with each annual acquisition (Figure 62). Therefore, an script was written in R (R Development Core Team 2009) to derive an extent that was common for each path/row/scene combination across years. This involved extracting the easting and northing value for the lower left and upper right corner of each grid in each scene from 2000 – 2011. The distance between every easting and northing combination was derived using the following formula:

$$Distance = \frac{\sqrt{(easting)^2 + (northing)^2}}{100}$$

The combination of easting and northing that was the closest was then used to trim each scene to derive a common extent.

A large table with common extents was created containing each unique easting and northing, and the ground cover percentage value for each year. Each scene contained null values (identified by a value of -1) representing the masked attributes and missing values in the grids. To ensure a consistent set of data was used between scenes all of the null values were summed and any cells that had a sum of -11 (all years with null values) where removed.

Figure 62. An example of a Landsat TM scene visualising the extent of each year stacked to illustrate the variation in extent and null values. The common extent is outlined in red.

A linear model and mean were derived using R (R Development Core Team 2009) for each scene following the methods outlined in Wallace *et al.* (2004). Due to the large size of each data frame required to run the linear model (~3.2 gigabytes) it was necessary to optimise the script and utilise four, four-gigabyte nodes on the JCU HPC for each scene (48 nodes representing 576 gigabytes of memory used simultaneously). Even with 12 gigabytes of dedicated memory each linear model took over 30 hours to process and a further 10 hours to derive a mean value per pixel in every scene (a total of 480 hours or 20 days of processing time).

Derived values were based on the percentage of each pixel representing bare ground and were expressed in a value between zero and one. Mean percentage bare ground was derived for each pixel in the Pilbara bioregion using the 10 years of cover values. The linear model estimates a value for each pixel representing no change (a value of zero) increasing bare ground (a gradient of positive values with higher values representing a consistent increase in bare ground over the 10 year period) and decreasing bare ground (a gradient of negative values with lower values representing the greatest increase in cover over the 10 year period).

6.8 Results

Linear models were developed for all fractional ground cover scenes in the Pilbara region. Application of these models to biodiversity habitat condition assessment, however, requires additional contextual analysis before they can be applied across the region. The mean percentage ground cover for the Pilbara region (Figure 63) clearly highlights the areas with naturally sparse vegetation such as rocky hills and areas of high cover. To adequately interpret the data a land classification map representative of the model resolution

and inherent environmental heterogeneity (30m resolution) is required. This would be used to normalise the average condition between land classes. Taking also into account temporal rainfall data, a normalised change index would highlight areas that are above or below the expected ground/vegetation cover due to human-mediated disturbance processes. Once the deviance from the expected ground cover is derived the coefficient value from the linear model (Figure 64) can be used to identify a variety of functional states. For example, areas that are degraded and not recovering will have a low percentage ground cover and a coefficient value close to zero whereas areas that are recovering will have a low percentage cover and a positive coefficient value greater than zero.
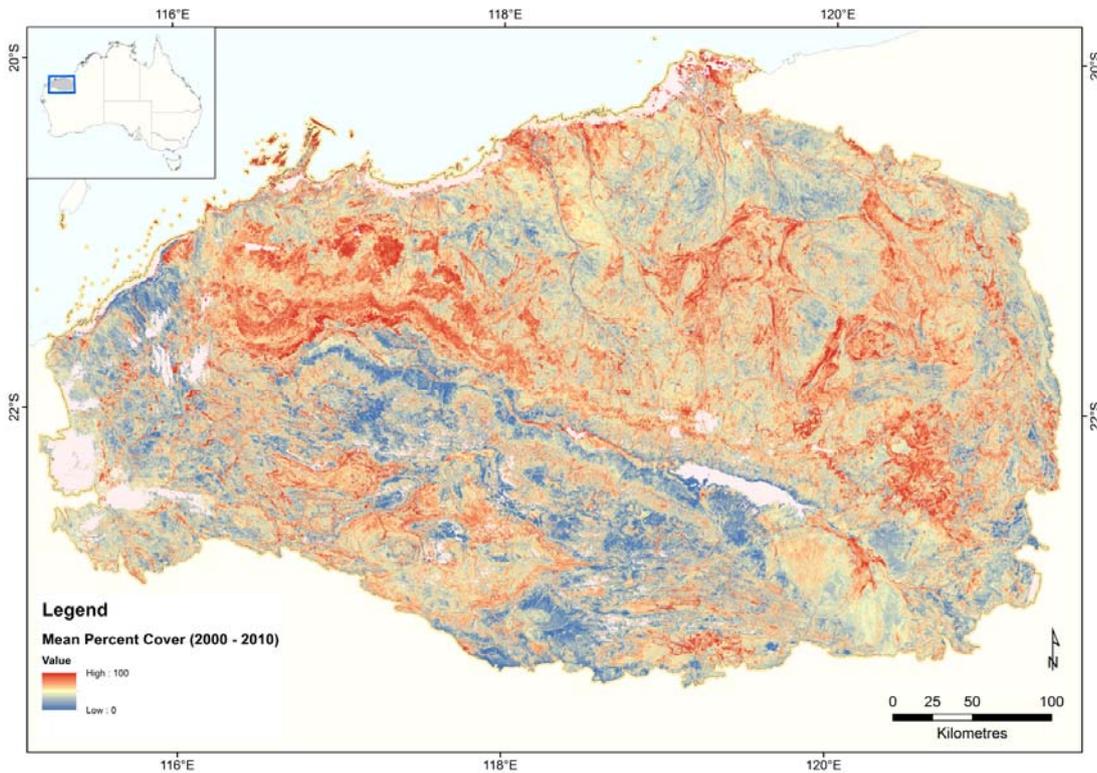
Figure 63. Regional mean percentage ground cover between 2000 and 2010. Note the areas with gradients tending towards dark red highlight rocky areas and areas that are naturally more sparsely vegetated but also highlight infrastructure. This highlights the potential improvements that can be achieved if normalising cover variation using a classification of the inherent environmental heterogeneity.
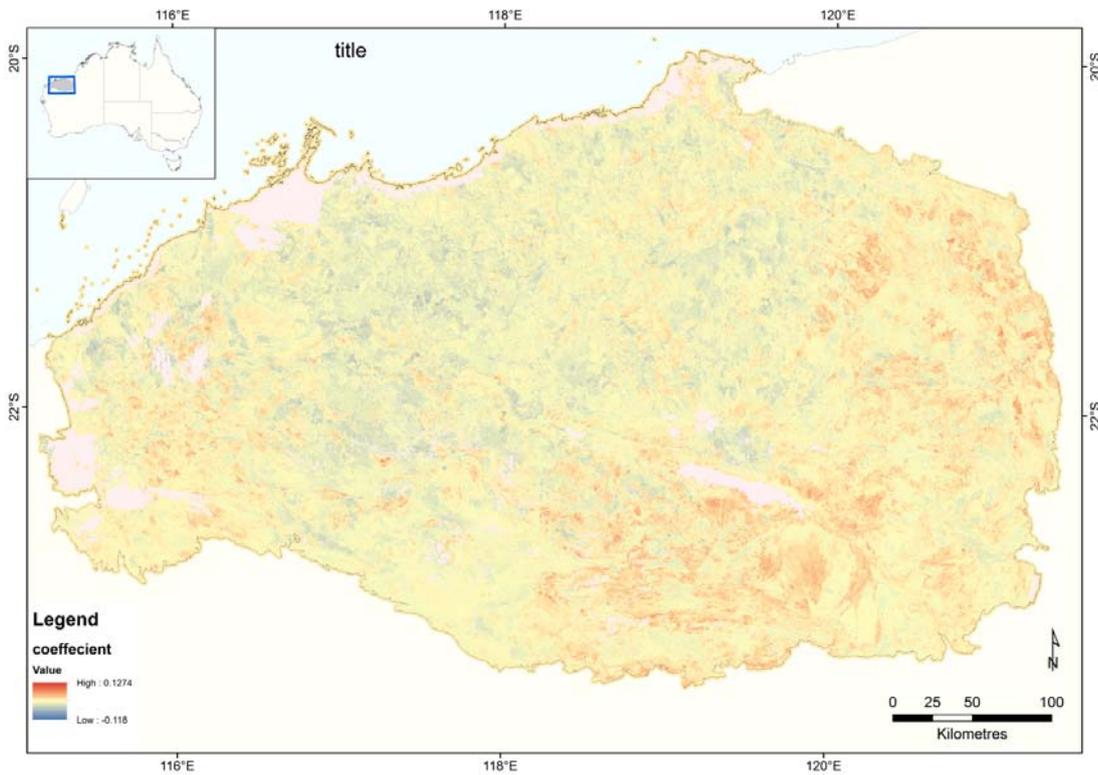
Figure 64. Regional linear model illustrating positive (gradient to dark red), negative (gradient to blue) and more stable (yellow) ground cover in the Pilbara region between 2000 and 2010.

This method can accurately distinguish the heterogeneity of ground cover mediated by infrastructure and different management regimes. In the context of the mining industry, this method can be used to identify the extent and impact of operations (Figure 65) and monitor the changes in cover following remediation and rehabilitation works (Figure 66). In a broader sense Figure 67 illustrates the distinct boundaries of vegetation cover influenced by the presence of a primary road and paddock boundaries. The area south of the road is inside a reserve and the area to the north is pastoral tenure. It is clear that the management strategies to the north of the road are significantly different to that south of the road. This is apparent as a high coefficient in the linear model suggesting continuing reduction in ground cover over the past 10 years.
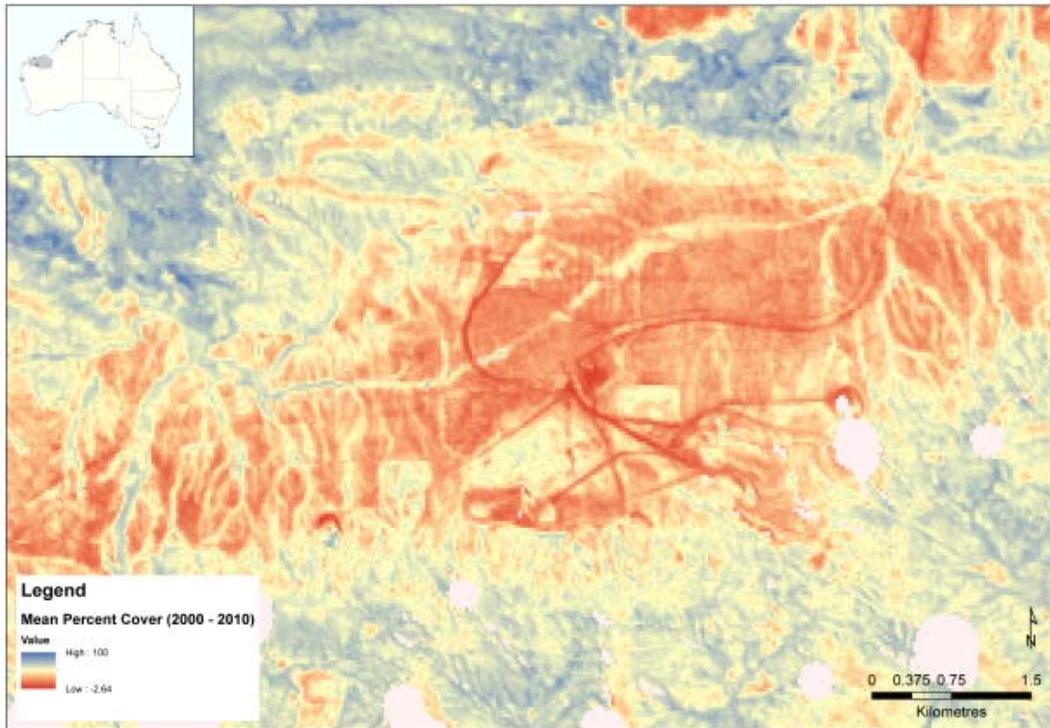
Figure 65. Mean percentage cover for a large mining area in the Pilbara region, clearly illustrating the extent of the operation.
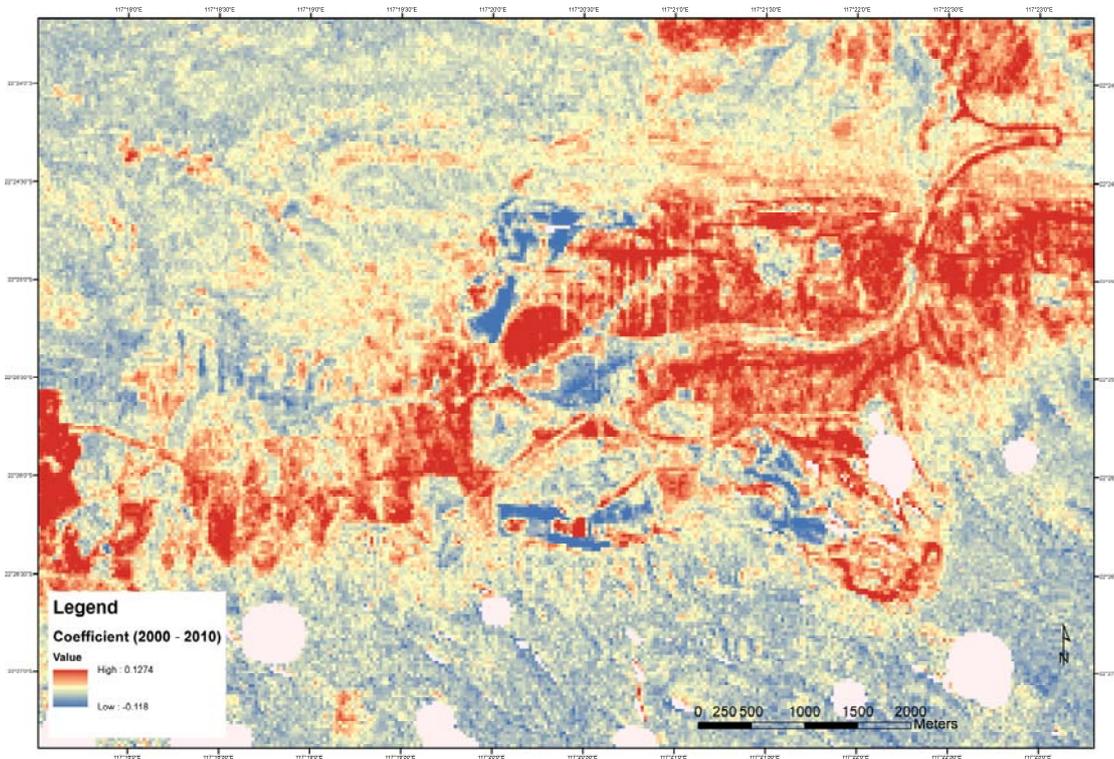
Figure 66. Displaying the coefficient from the linear model at the same location depicted in the figure 30 illustrating areas that have increased in cover (tending to blue) or decreased in cover (tending to red).
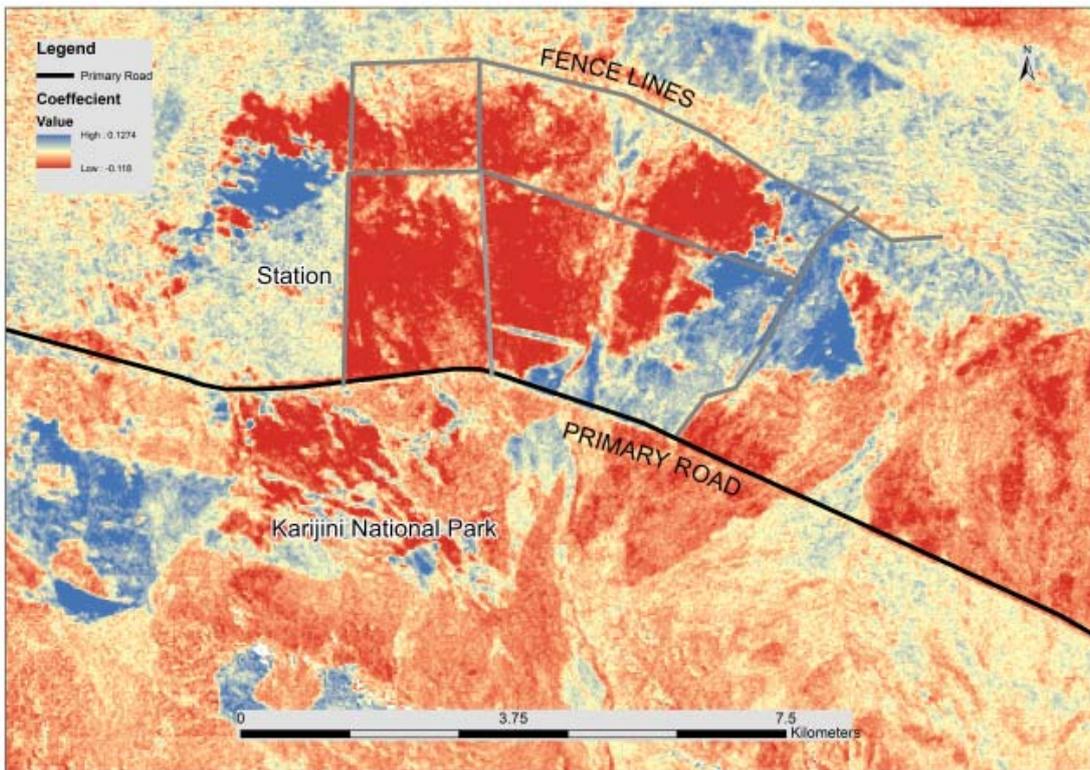
Figure 67. Example linear model output depicting areas where vegetation cover has changed negatively (red) and positively (blue) or remained constant (yellow) over the past 10 years based on 30m pixel resolution. The relative impact of infrastructure (roads and fence lines) and varying management (grazing and reserved land) can be clearly seen north and south of the road.

Appendix F Biodiversity significance maps

Summary

The maps in this appendix summarise the various components of the biodiversity significance analyses. All maps have been masked to the Pilbara bioregion. Results are presented for each group, starting with All Groups, followed by Birds, Invertebrates, Mammals, Vascular Plants and Reptiles. The calculations in each case are given in Section 5.2. For each group, where appropriate, the following maps are presented:

a) Biodiversity significance excluding richness   i) no condition (Equation 1)

ii) regional condition (Equation 2)

iii) regional and local condition (Equation 3)

b) Biodiversity significance including richness   i) no condition (Equation 4)

ii) regional condition (Equation 5)

iii) regional and local condition (Equation 6)

c) Richness model   i) normalised (contribution to biodiversity significance)

ii) absolute richness (see also Section 3.2)

d) Uncertainty surfaces   i) GDM survey coverage

ii) richness model uncertainty

Note that whilst the colour ramps used in maps within the four categories above are consistent, allowing the comparison of the different condition approaches, they may not be directly compared between categories. For example, b-i) and b-iii) have identical colour ramps but a-i) uses a different colour ramp. This may not be obvious, so checking of the legend is recommended.

It is also worth noting that the different analyses are expected to yield different value ranges, and that the expression of these across different groups with varying rates of compositional turnover may also vary. The difference in values across space *within* each map is the key feature, whereas apparent differences between maps in absolute values may have limited relevance.

CONTENTS

List of Figures

All groups

Figure 1 Biodiversity significance excluding richness and condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
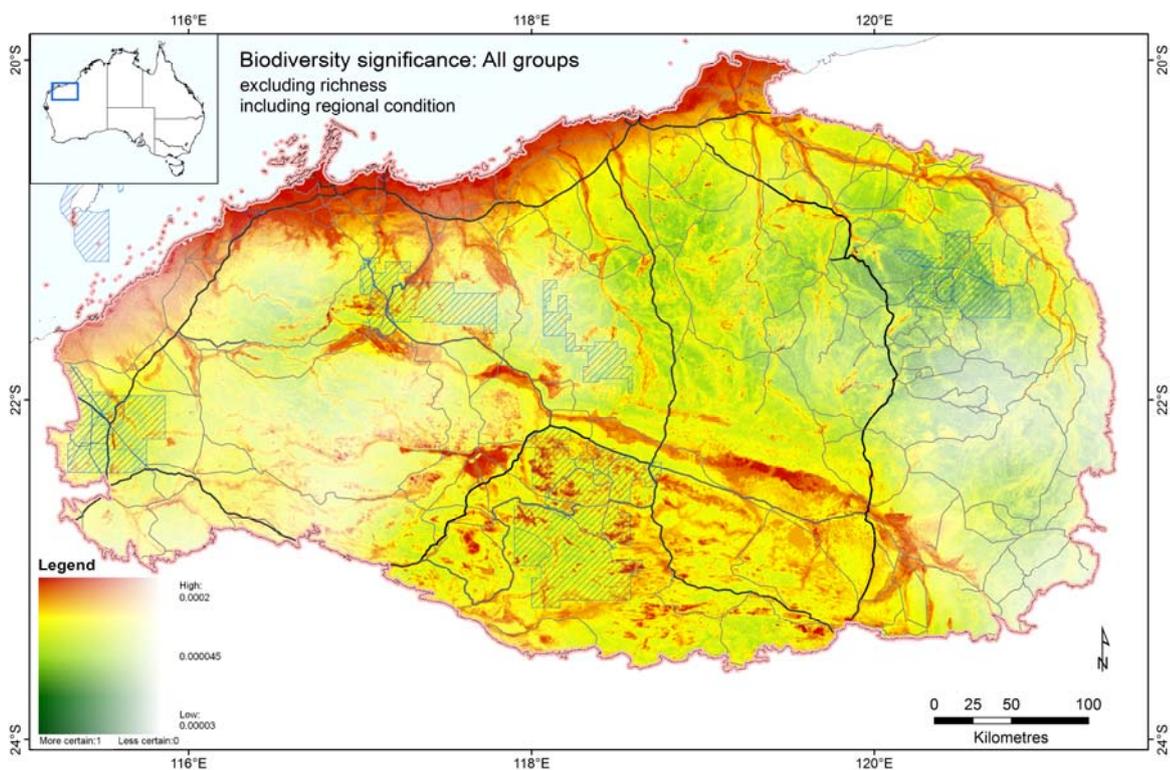
Figure 2 Biodiversity significance excluding richness and including regional condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
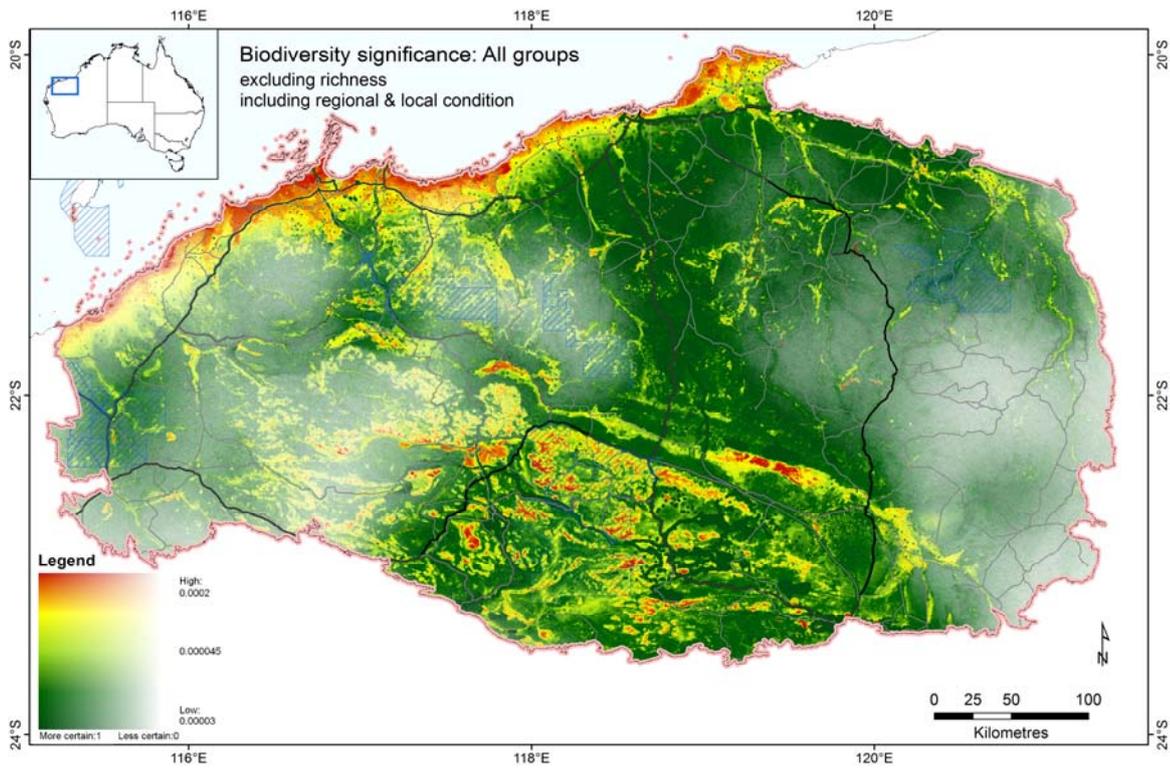
Figure 3 Biodiversity significance excluding richness and including regional and local condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
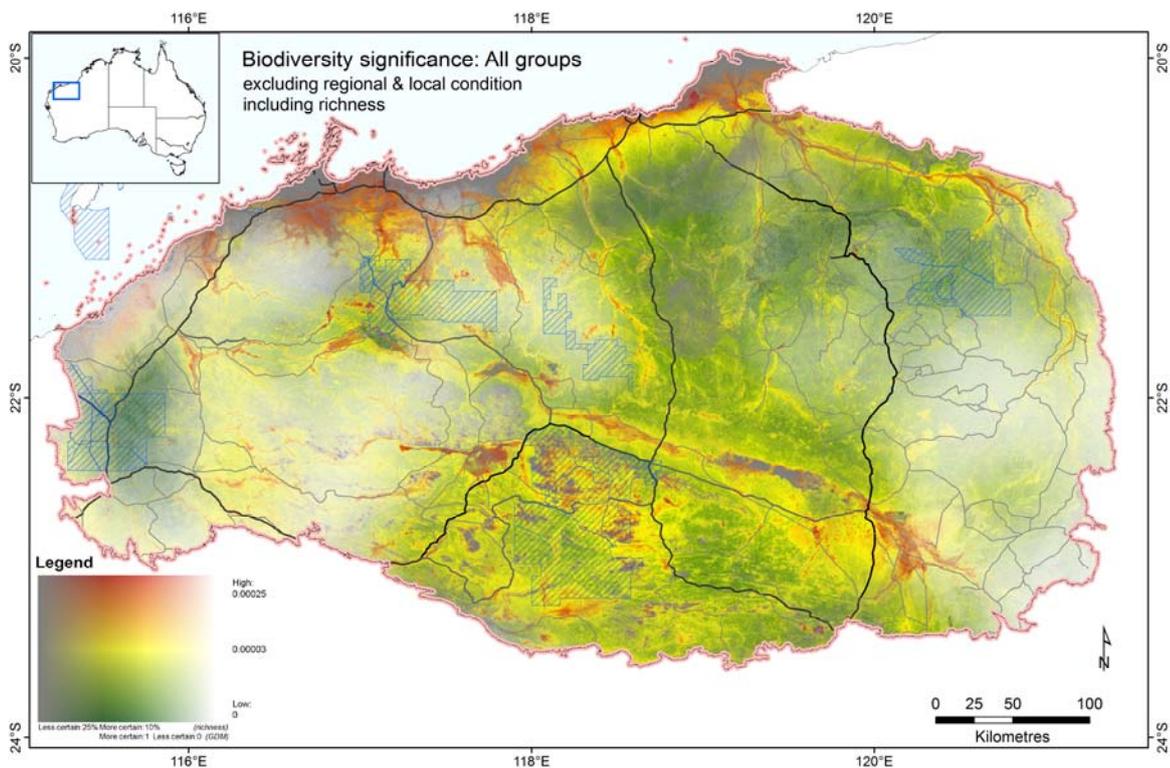
Figure 4 Biodiversity significance including richness excluding condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
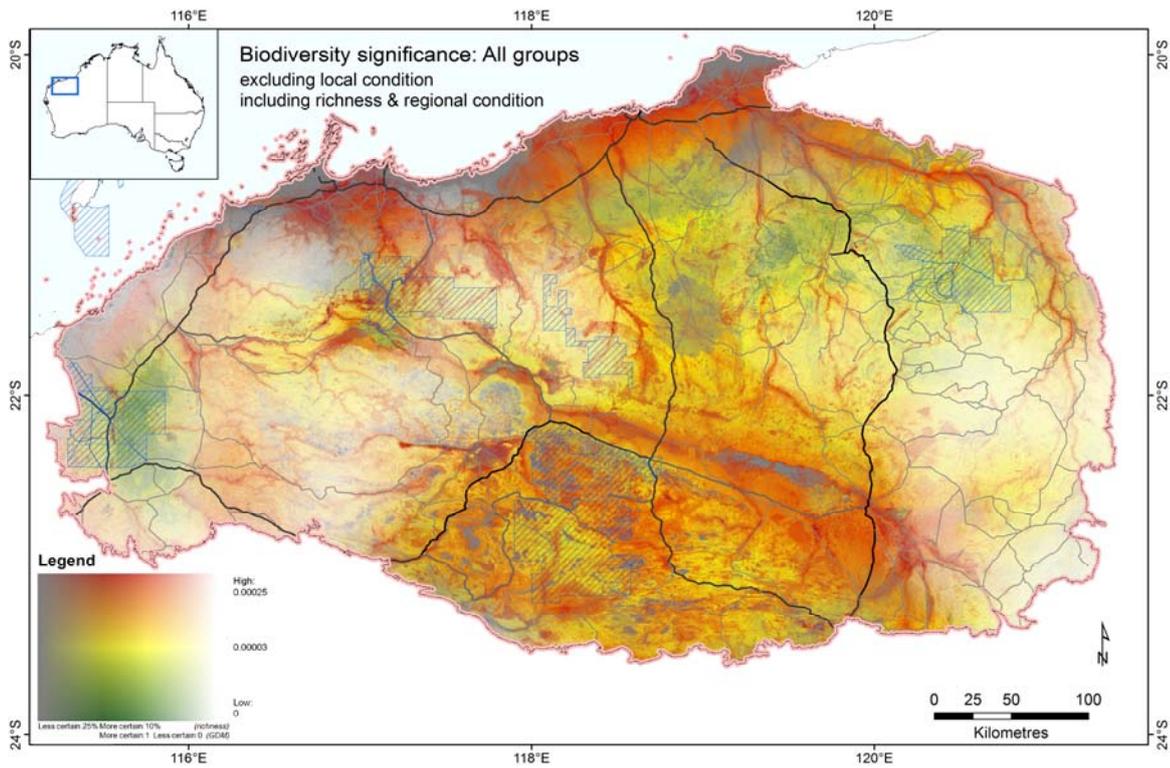
Figure 5 Biodiversity significance including richness and regional condition for all groups, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
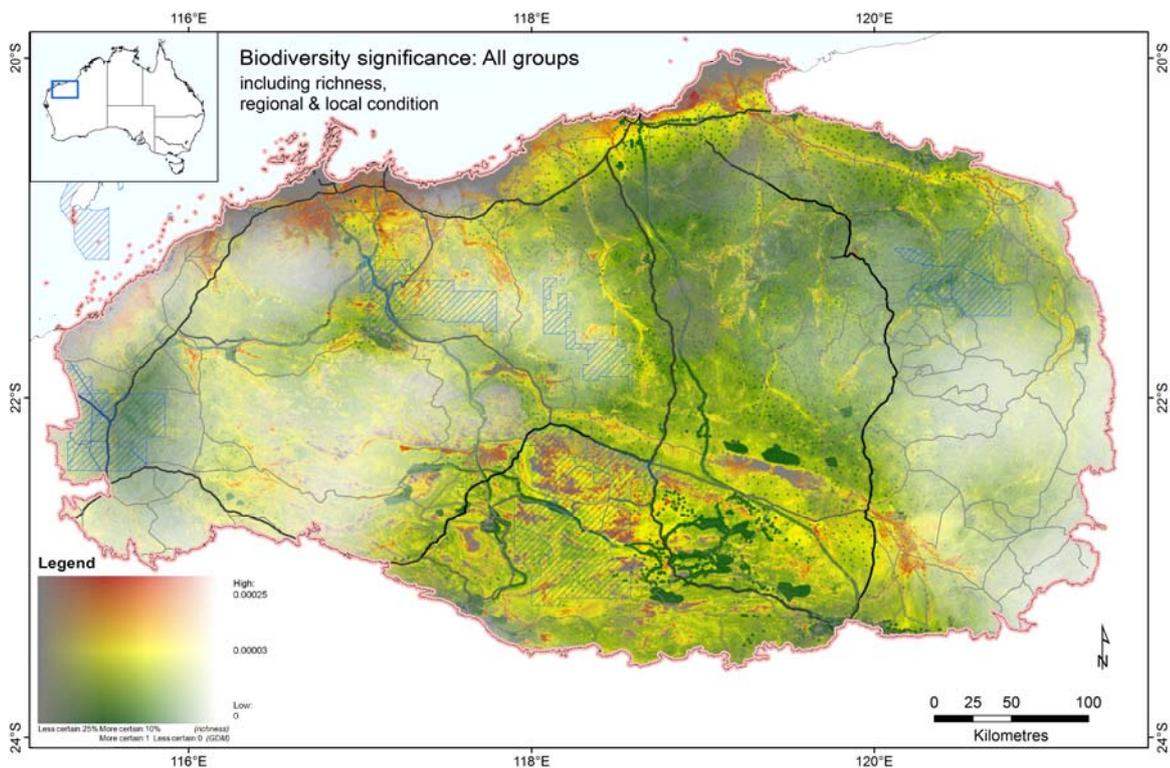
Figure 6 Biodiversity significance including richness, regional and local condition for all groups, based on community-level modelling. Significance is the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (GAM) areas more uncertain than transparent areas.
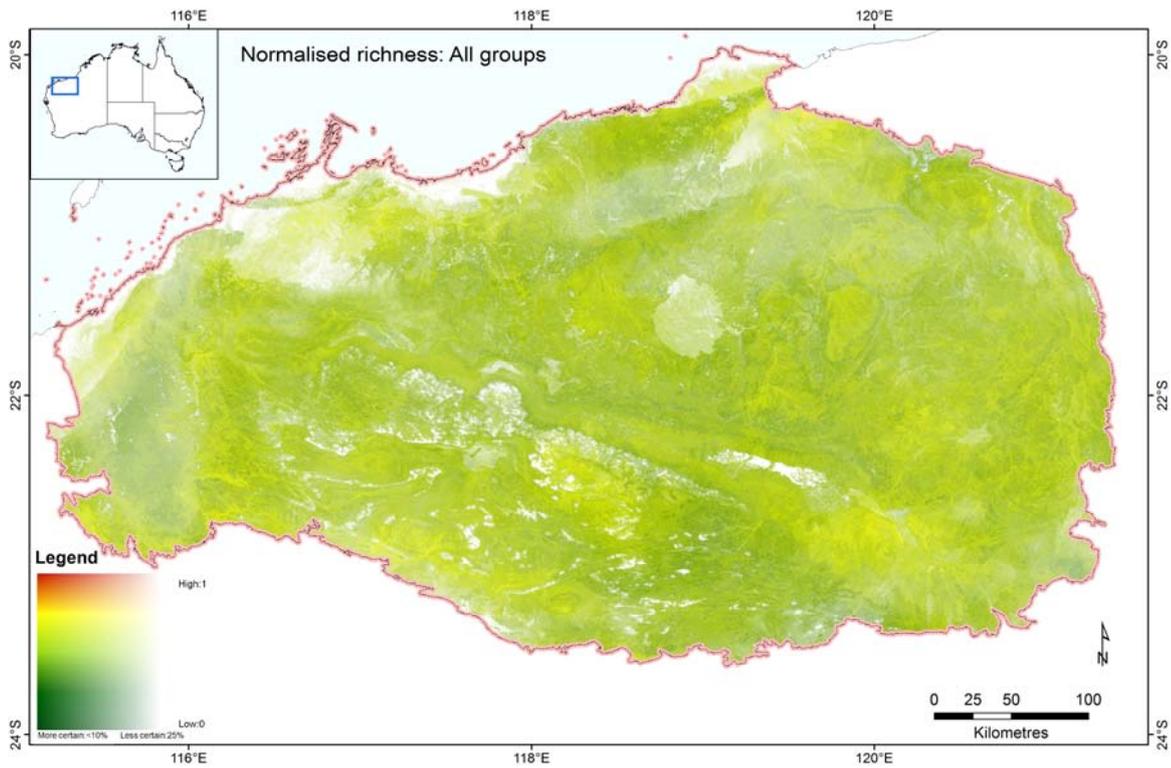
Figure 7 Normalised species richness (log fraction of maximum richness) for all groups, based on community-level modelling. This is the right hand side, ln(r_i)/ln(r_{max}) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.

Birds
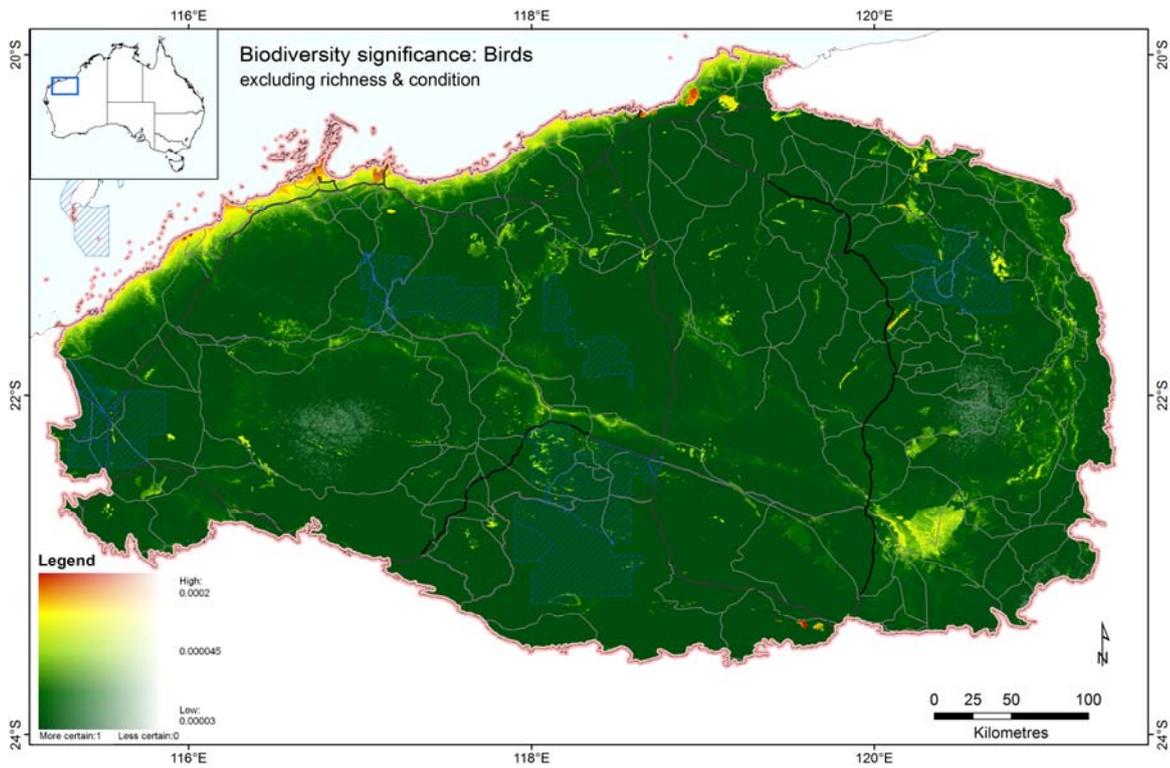
Figure 8 Biodiversity significance excluding richness and condition for Birds, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
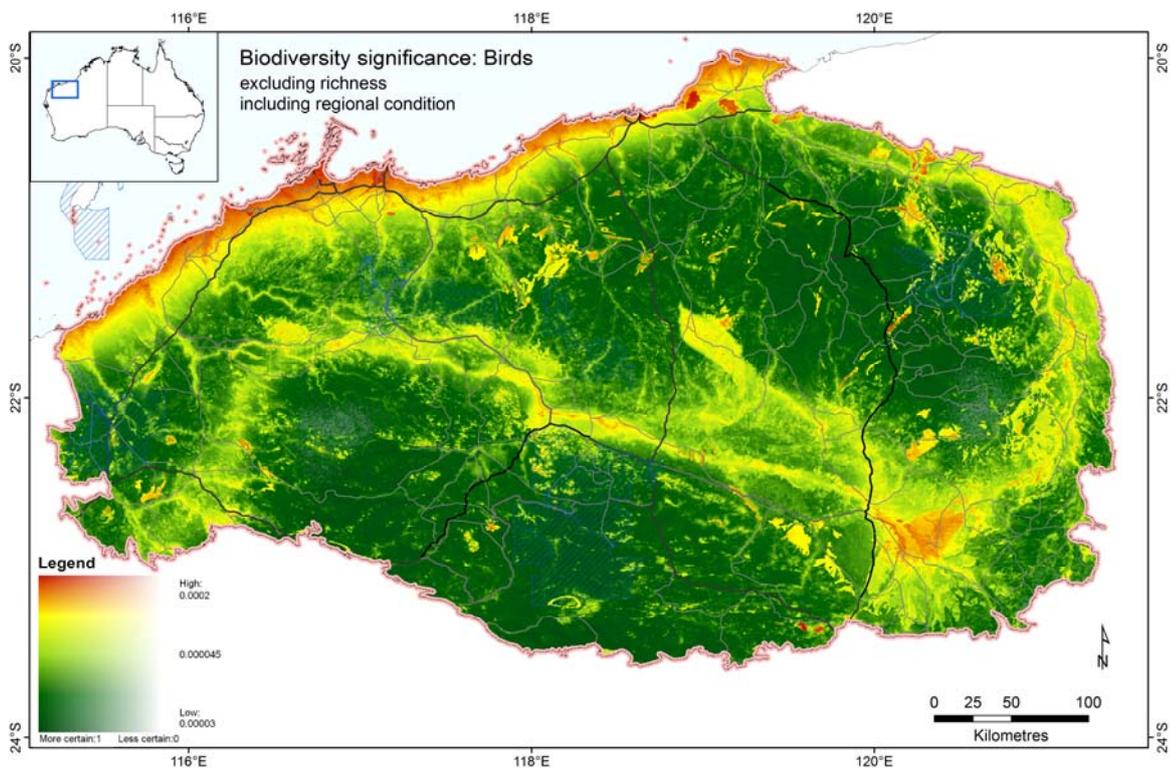
Figure 9 Biodiversity significance excluding richness and including regional condition for Birds, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
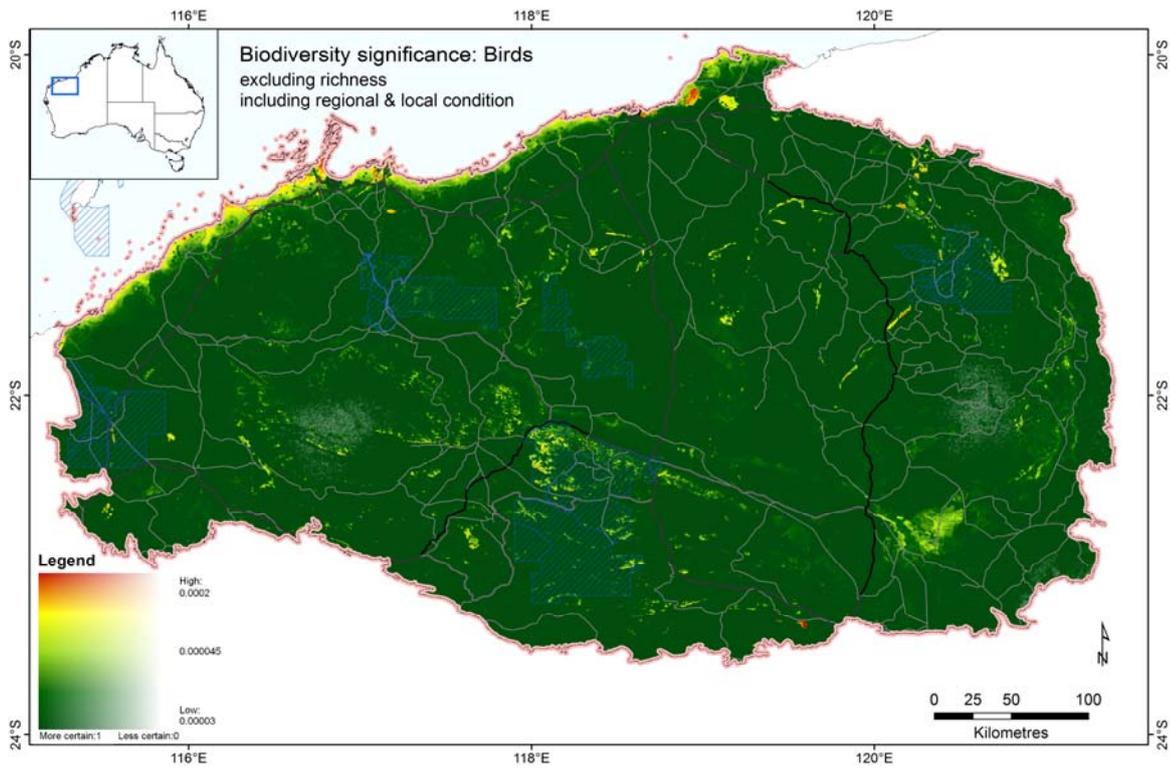
Figure 10 Biodiversity significance excluding richness and including regional and local condition for Birds, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
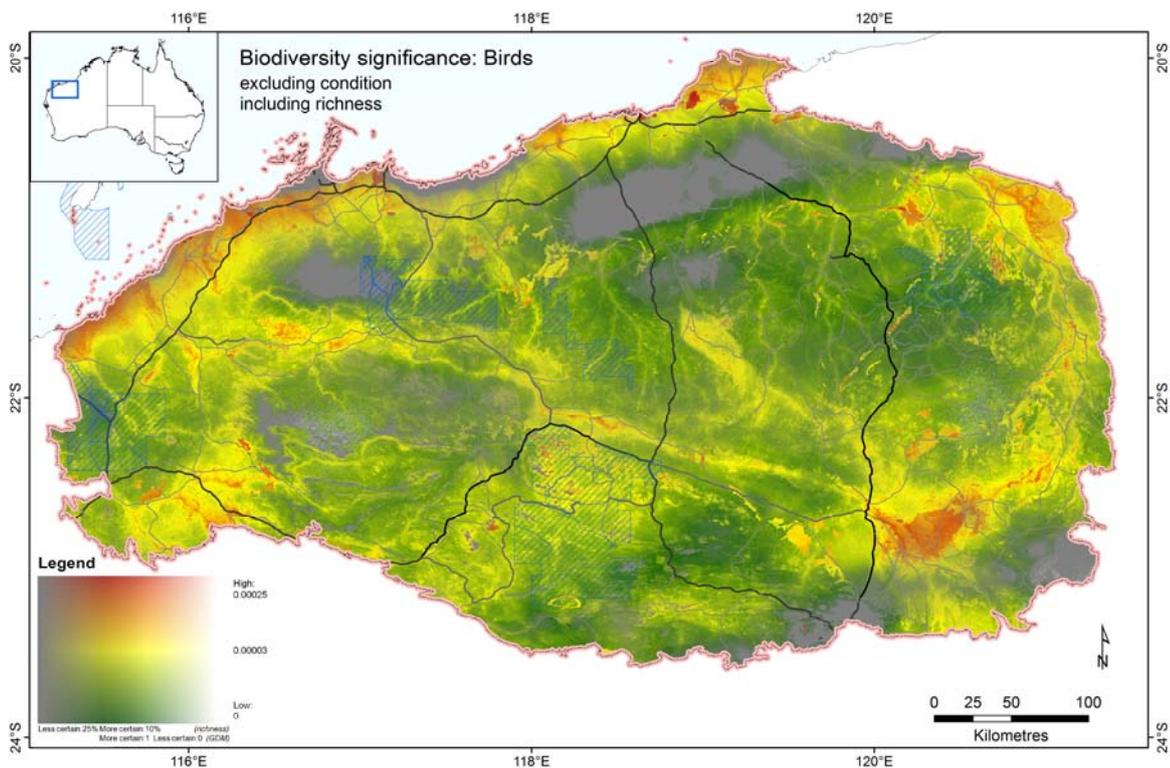
Figure 11 Biodiversity significance including richness excluding condition for all Birds, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
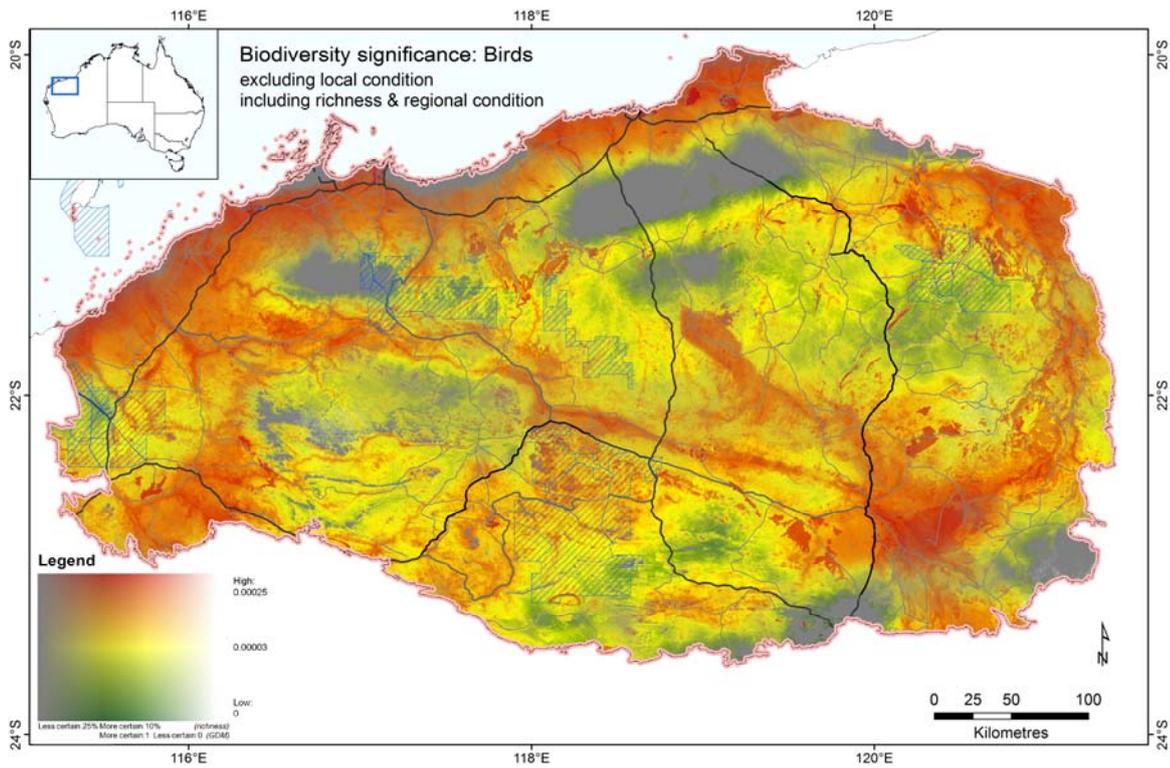
Figure 12 Biodiversity significance including richness and regional condition for Birds, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
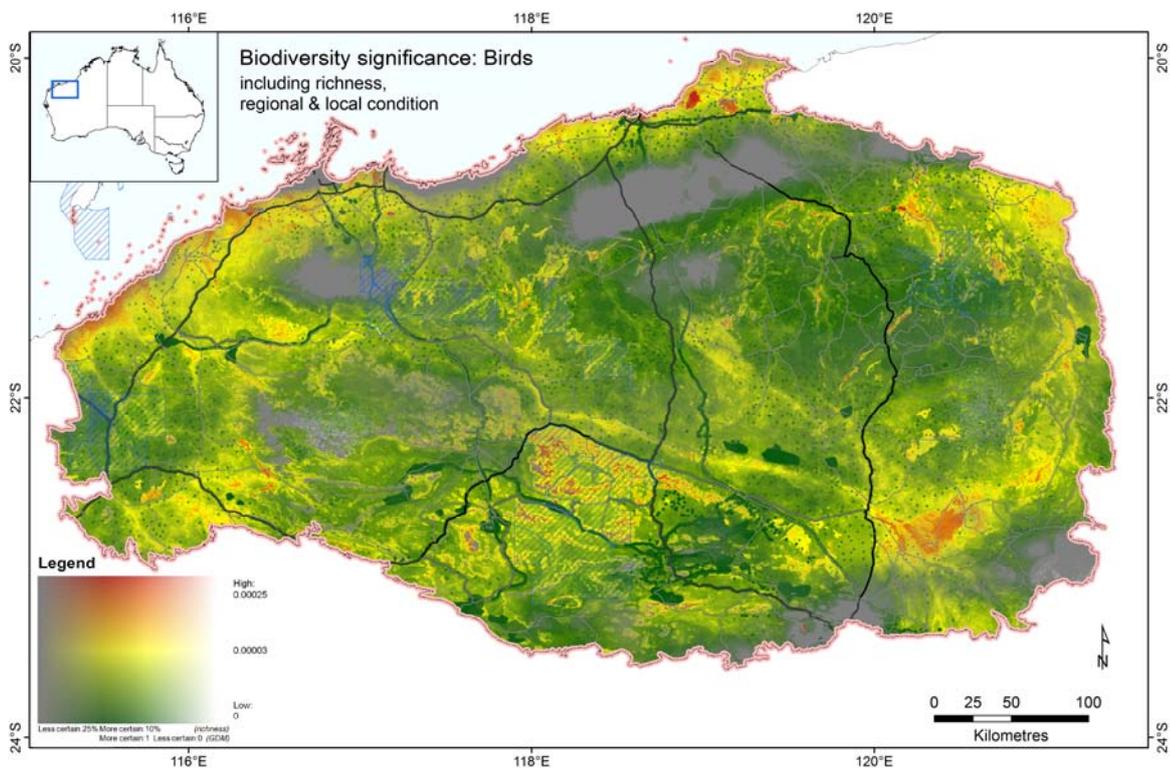
Figure 13 Biodiversity significance including richness, regional & local condition for Birds, based on community-level modelling. Significance calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
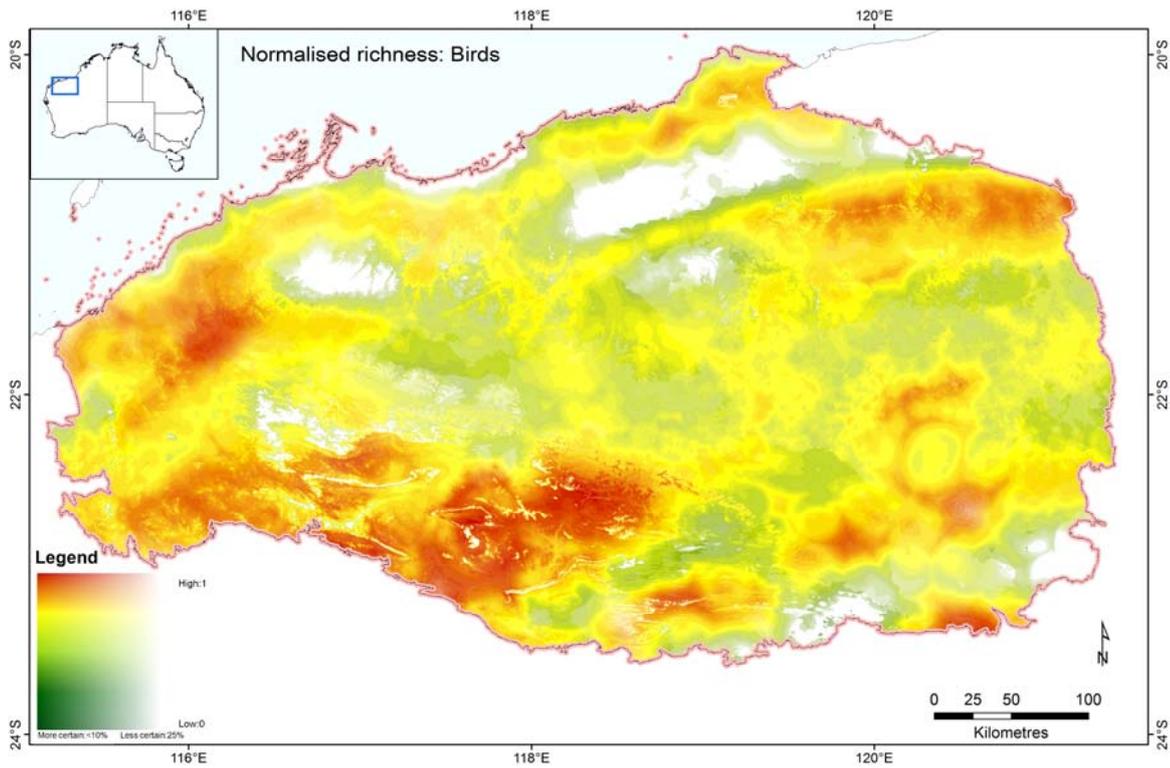
Figure 14 Normalised species richness (log fraction of maximum richness) for Birds, based on community-level modelling. This is the right hand side, ln(r)/ln(rmax) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
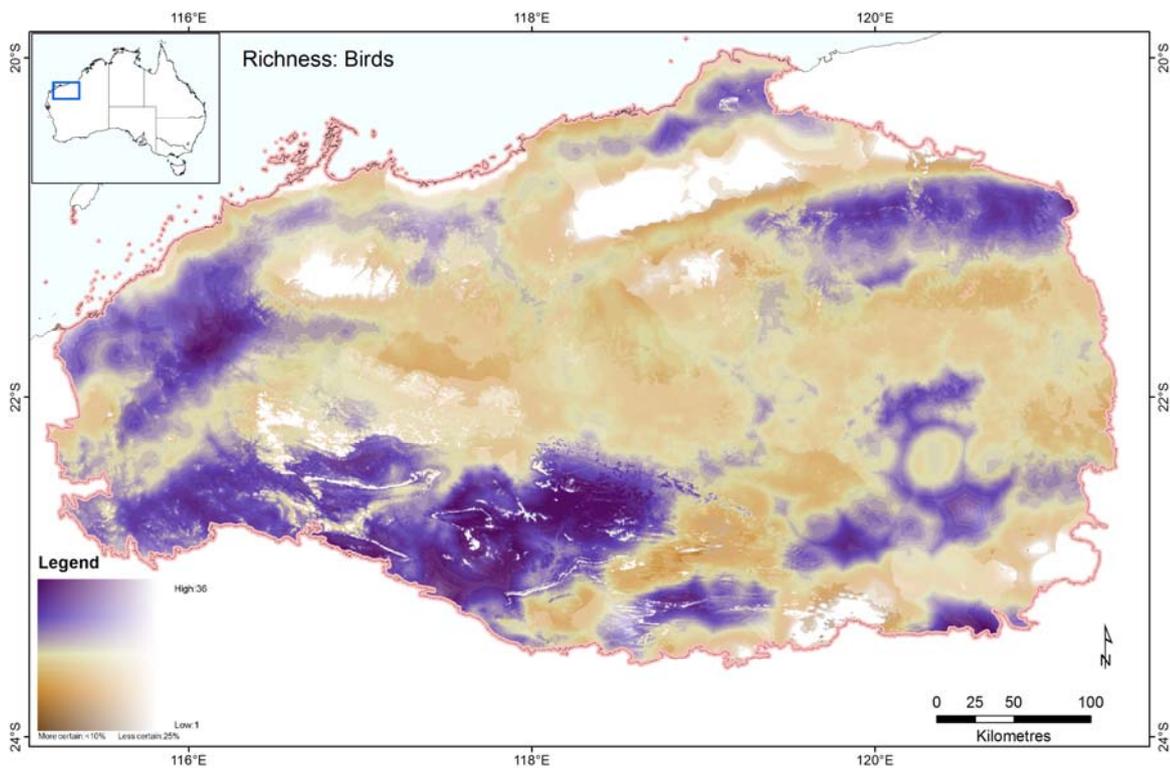
Figure 15 Absolute species richness (number of species in 9s grid cell) for Birds, based on community-level modelling. Purple areas have more species than brown areas. Whiter areas are more uncertain than transparent areas.
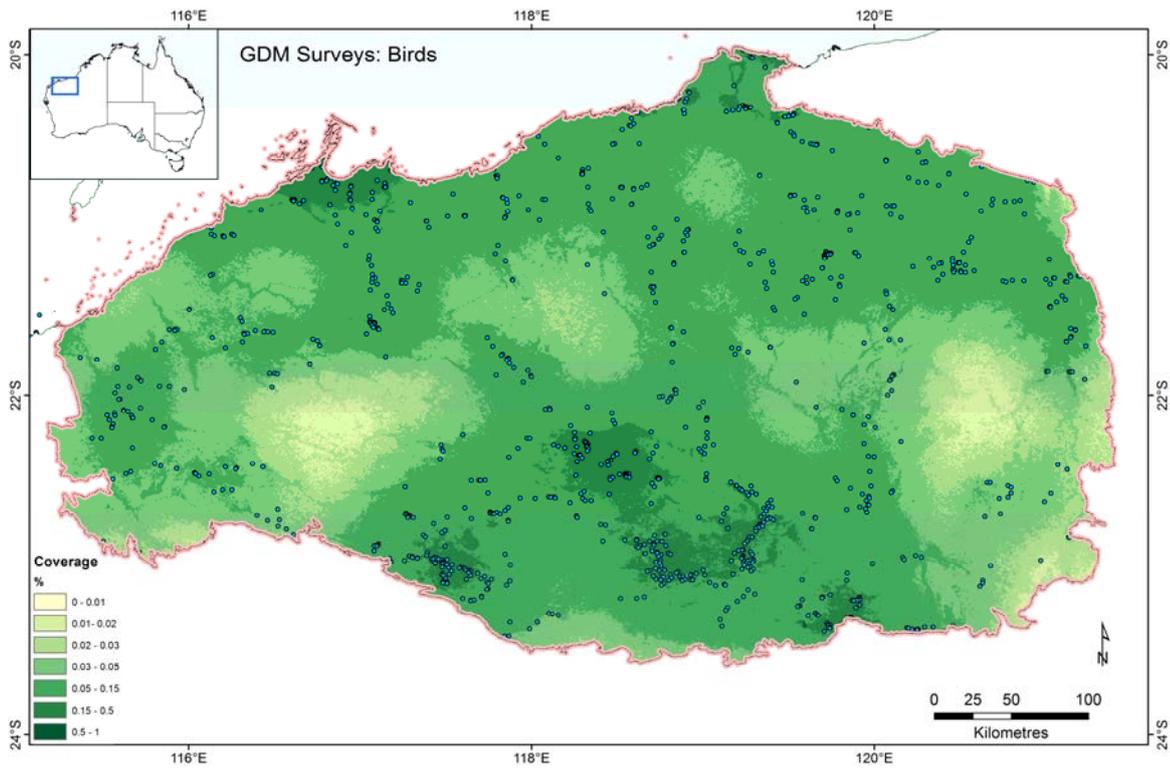
Figure 16. Location of surveys used in the Birds GDM analysis (blue circles) overlaid on a map of sample density in GDM scaled environmental space. This describes the proportion of similar habitat for each cell *i* which is covered by survey sites, ranging from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the GDM uncertainty cloud applied in subsequent maps.
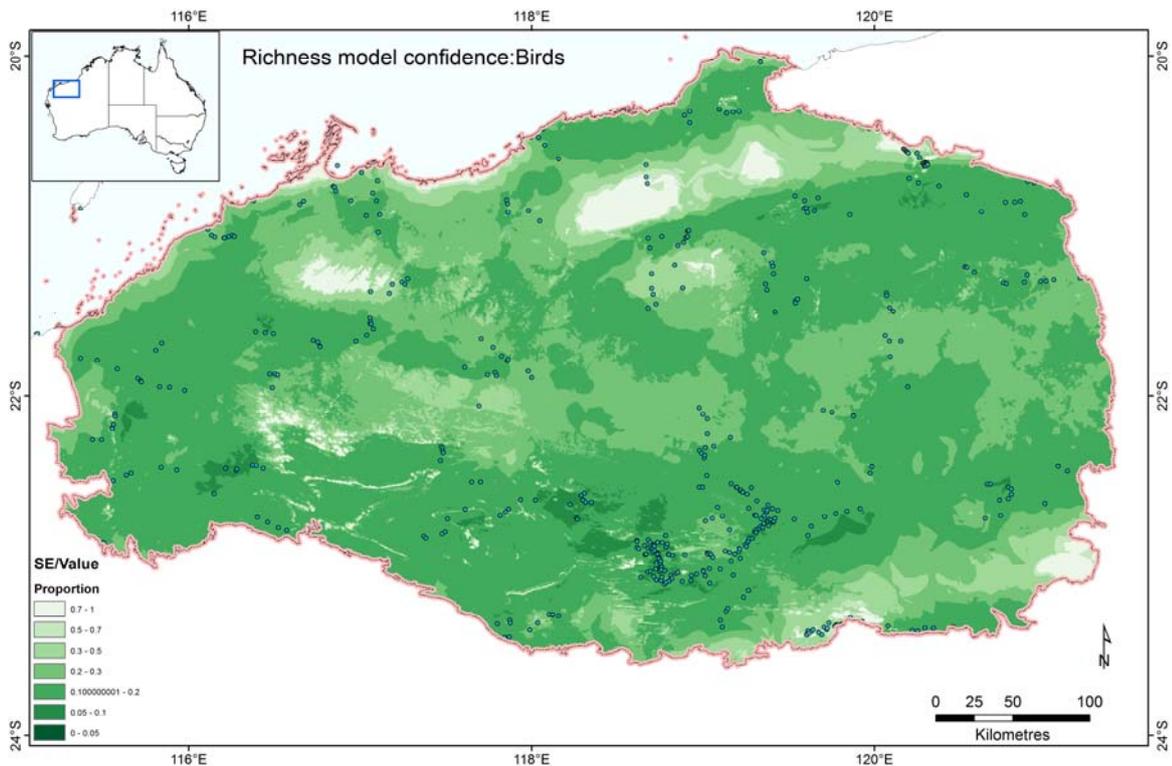
Figure 17. Location of surveys used in the Birds richness analysis (blue circles) overlaid on a map of richness model standard error/value. Values range from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the richness model uncertainty cloud applied in subsequent maps.

Invertebrates
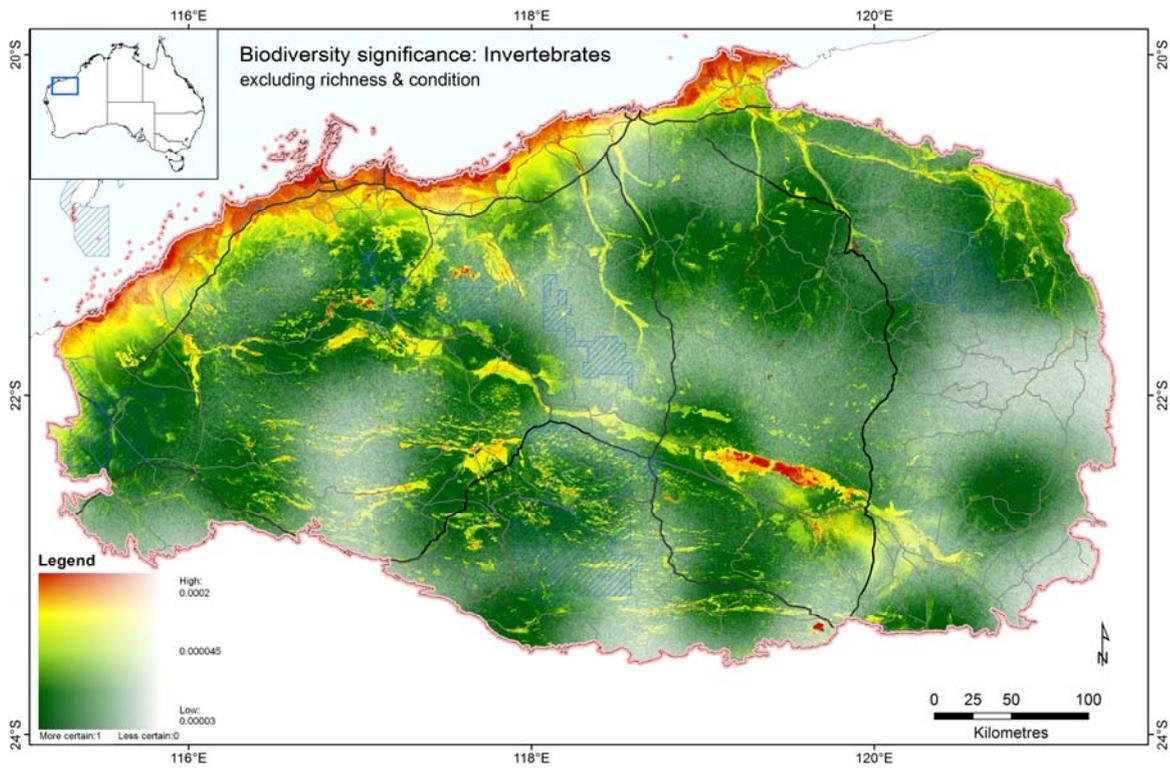
Figure 18 Biodiversity significance excluding richness and condition for Invertebrates, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
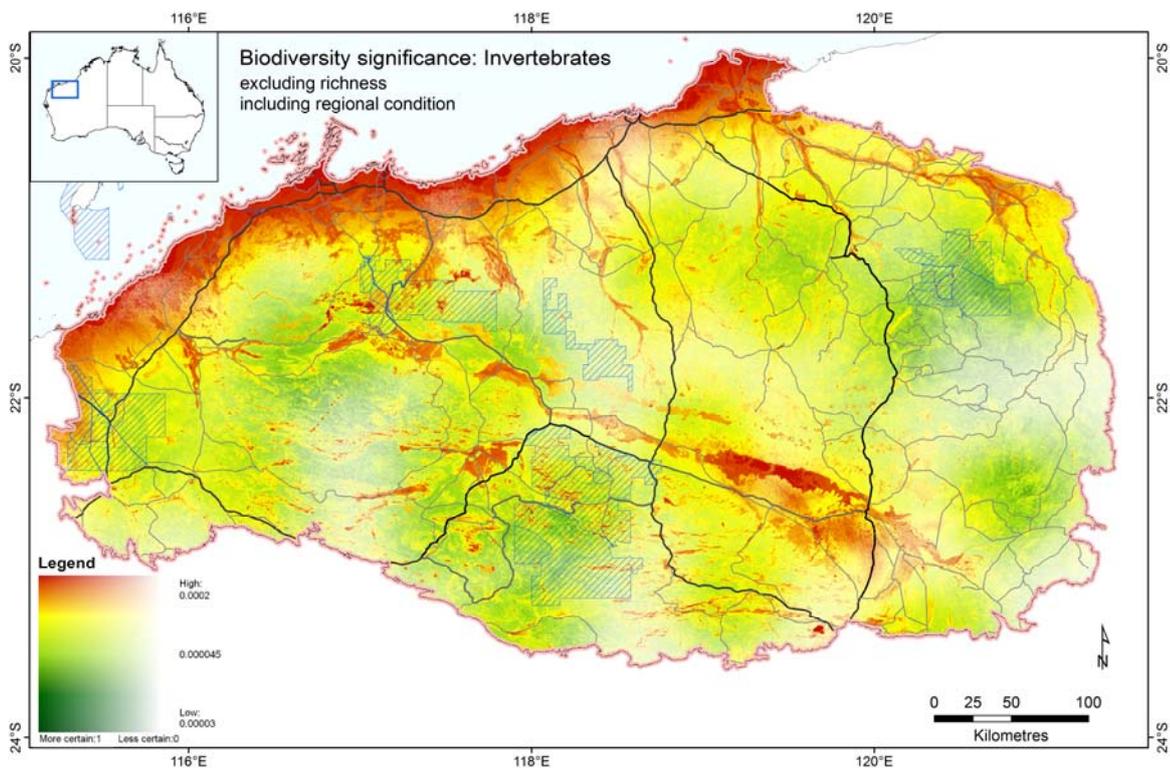
Figure 19 Biodiversity significance excluding richness and including regional condition for Invertebrates, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
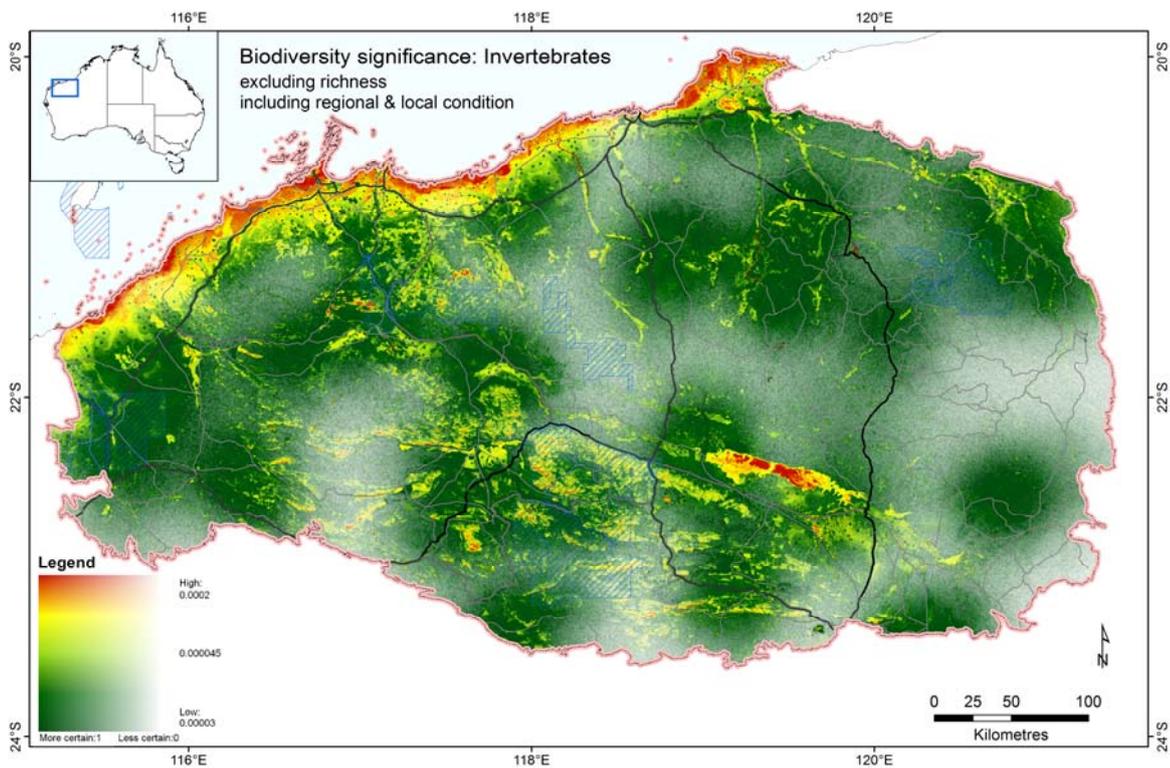
Figure 20 Biodiversity significance excluding richness and including regional and local condition for Invertebrates, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas more uncertain than transparent areas.
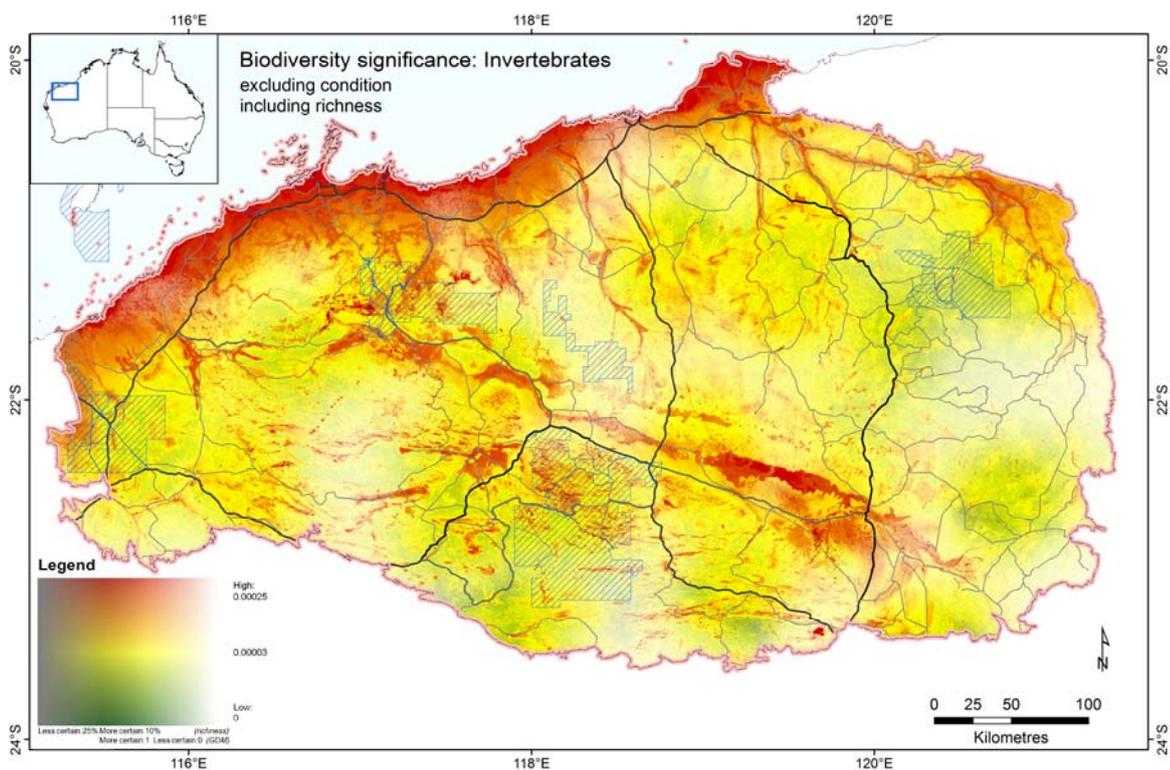
Figure 21 Biodiversity significance including richness excluding condition for Invertebrates, based on community-level modelling. Significance is calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
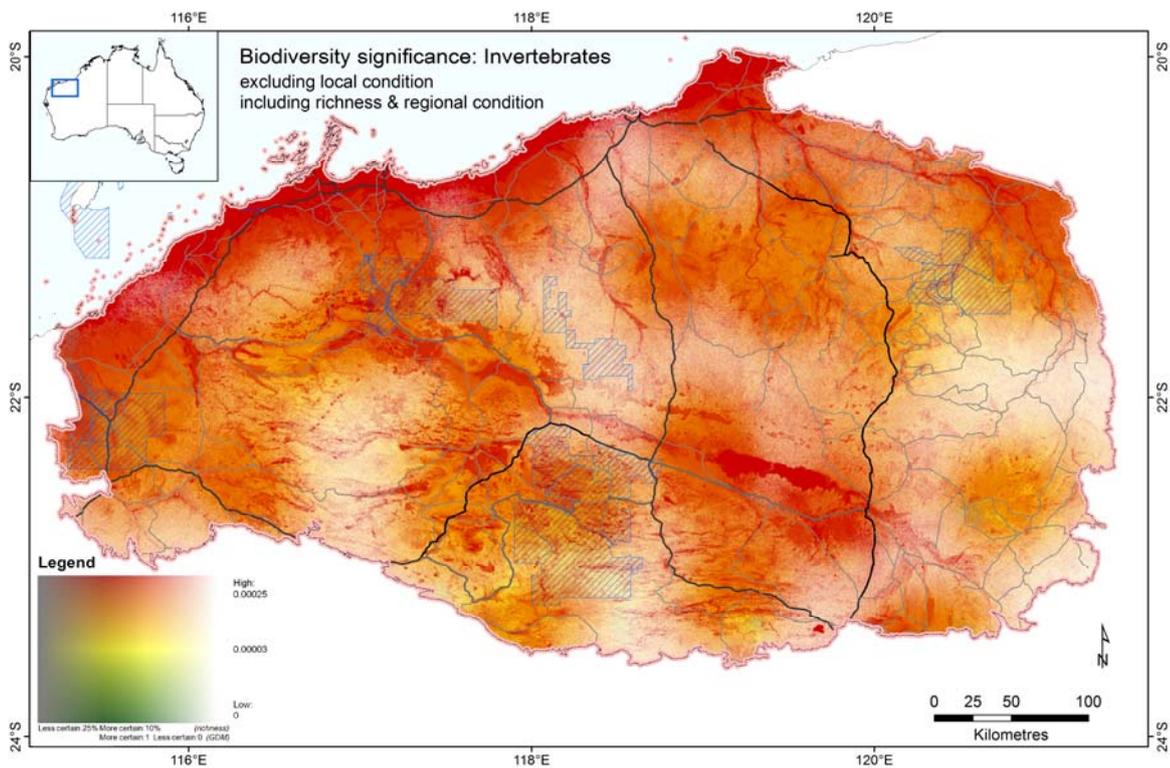
Figure 22 Biodiversity significance including richness and regional condition for Invertebrates, based on community-level modelling. Significance calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
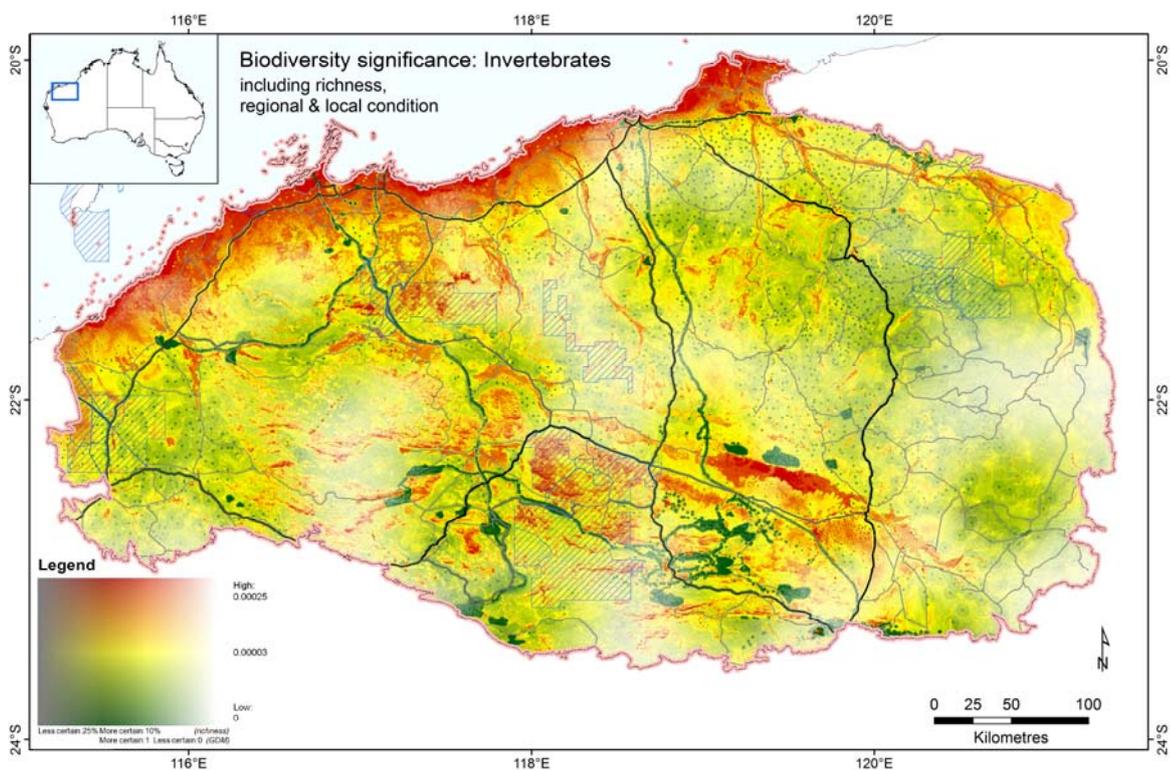
Figure 23 Biodiversity significance including richness, regional & local condition for Invertebrates, based on community-level modelling. Significance is the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
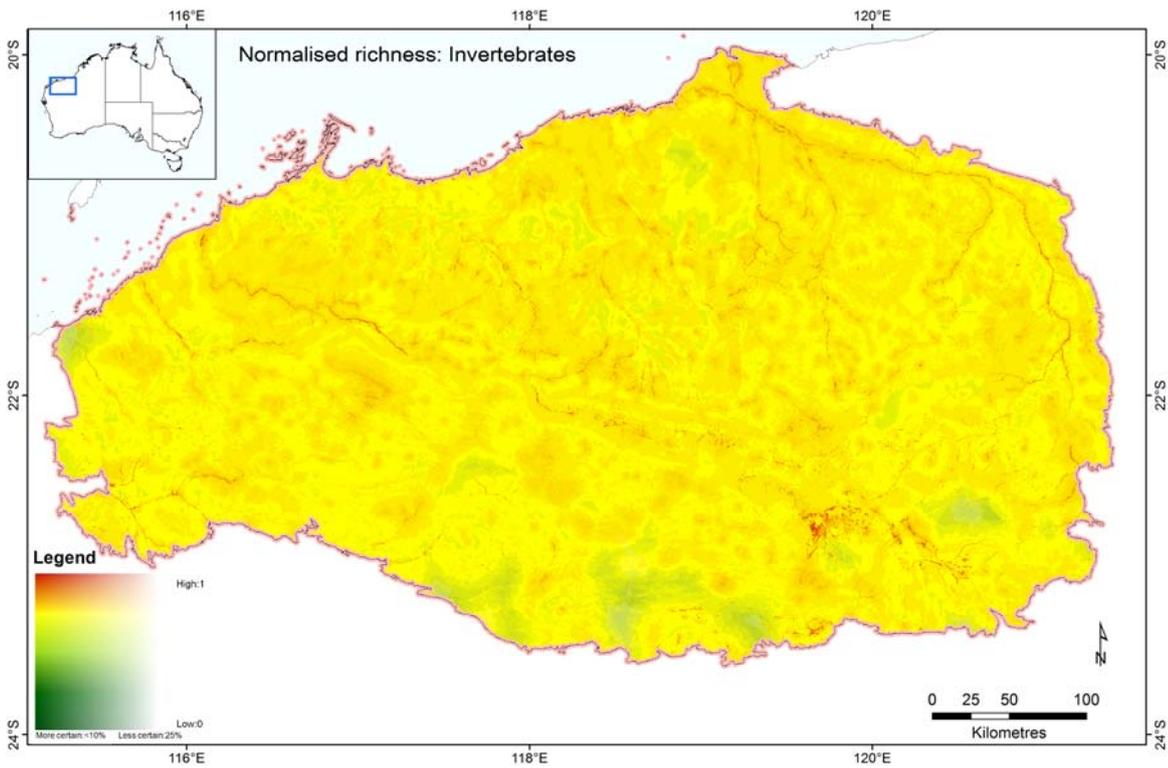
Figure 24 Normalised species richness (log fraction of maximum richness) for Invertebrates, based on community-level modelling. This is the right hand side, ln(r)/ln(rmax) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
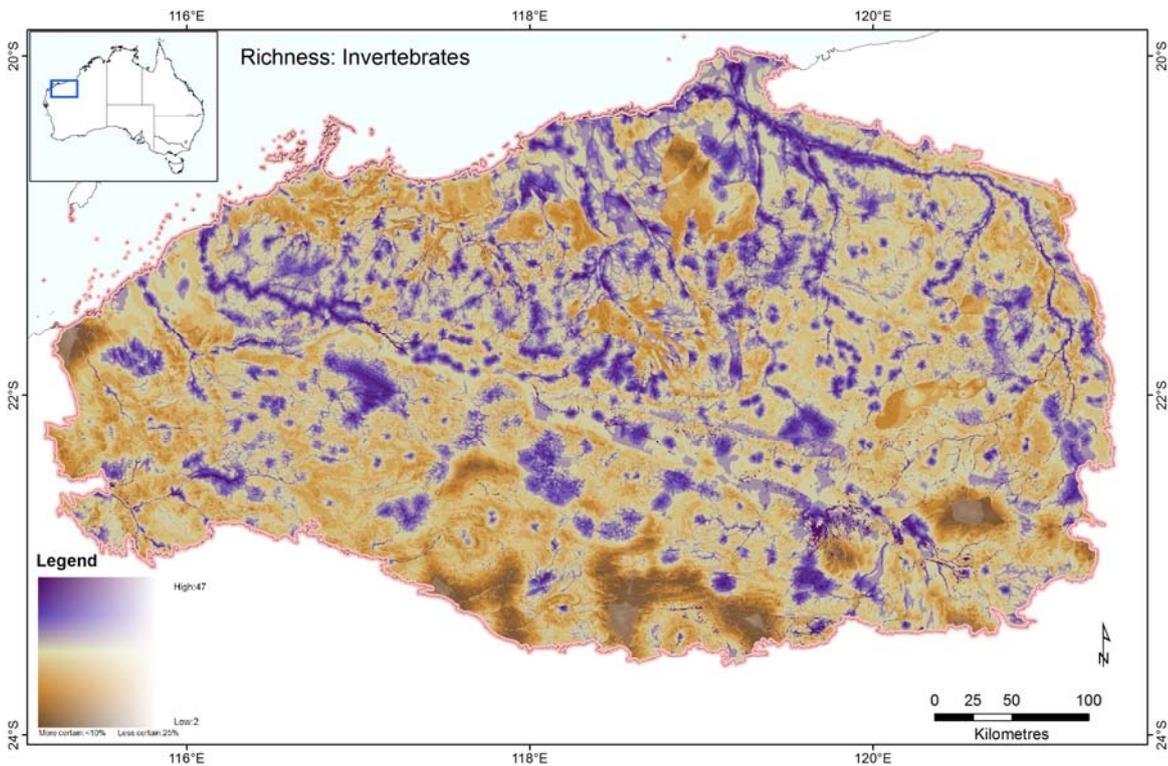
Figure 25 Absolute species richness (number of species in 9s grid cell) for Invertebrates, based on community-level modelling (see Section 3.2). Purple areas have more species than brown areas. Whiter areas are more uncertain than transparent areas.
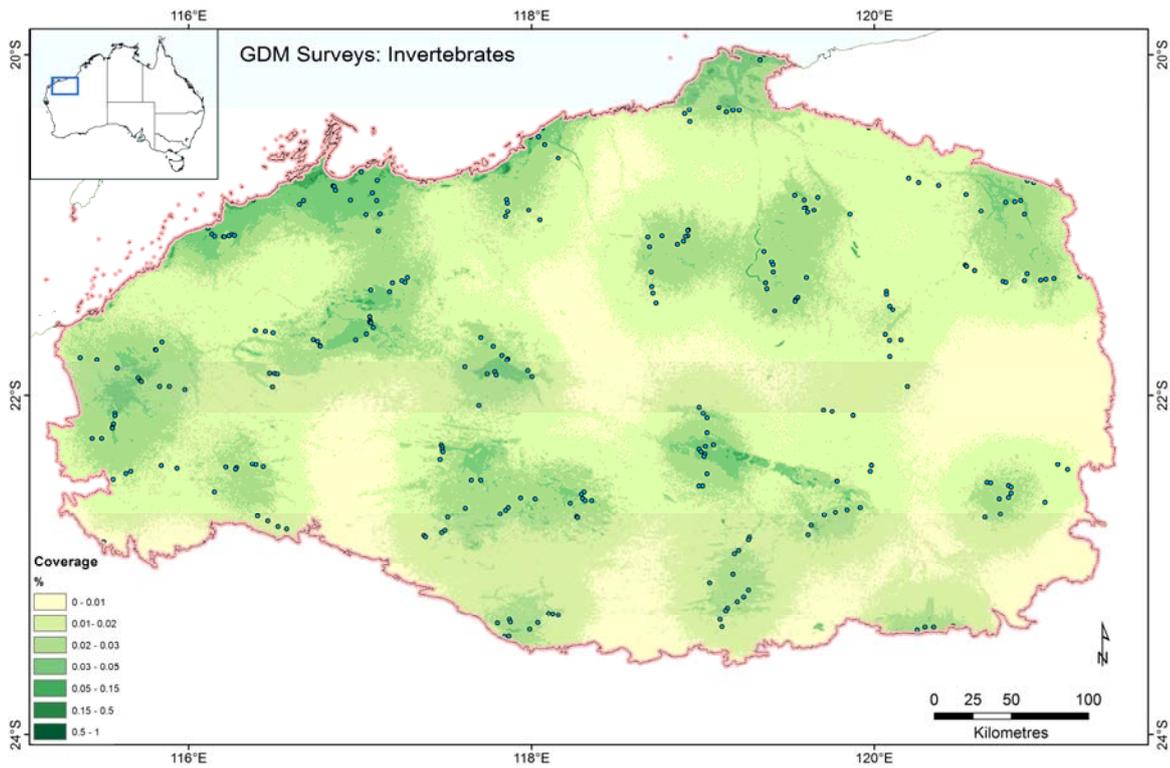
Figure 26 Location of surveys used in the Invertebrates GDM analysis (blue circles) overlaid on a map of sample density in GDM scaled environmental space. This describes the proportion of similar habitat for each cell i which is covered by survey sites, ranging from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the GDM uncertainty cloud applied in biodiversity significance maps.
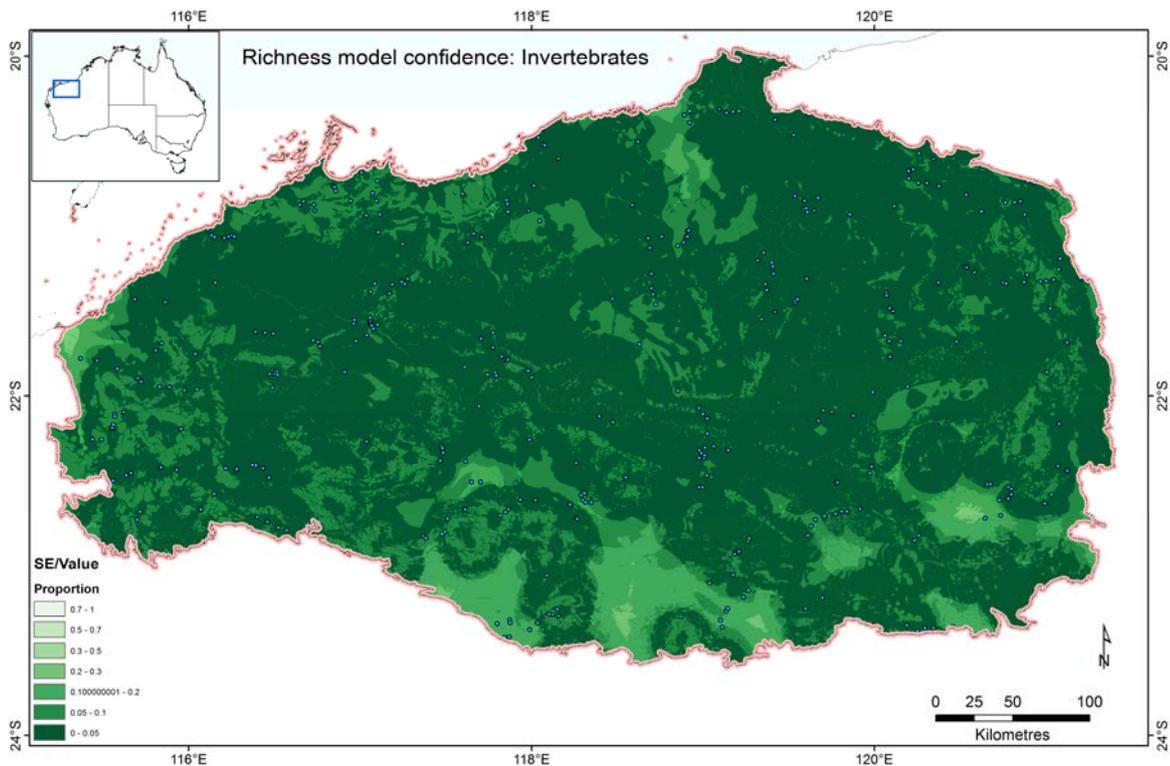
Figure 27 Location of surveys used in the Invertebrates richness analysis (blue circles) overlaid on a map of richness model standard error/value. Values range from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the richness model uncertainty cloud applied in biodiversity significance maps.

Mammals
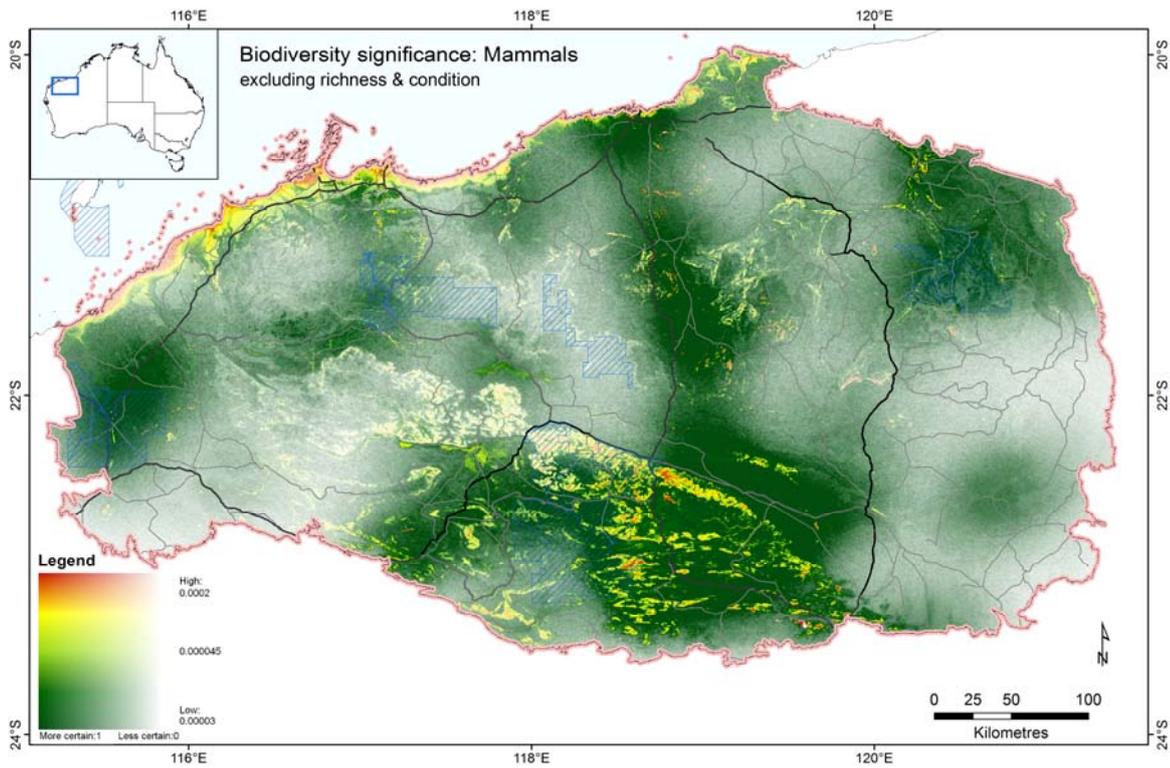
Figure 28 Biodiversity significance excluding richness and condition for Mammals, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
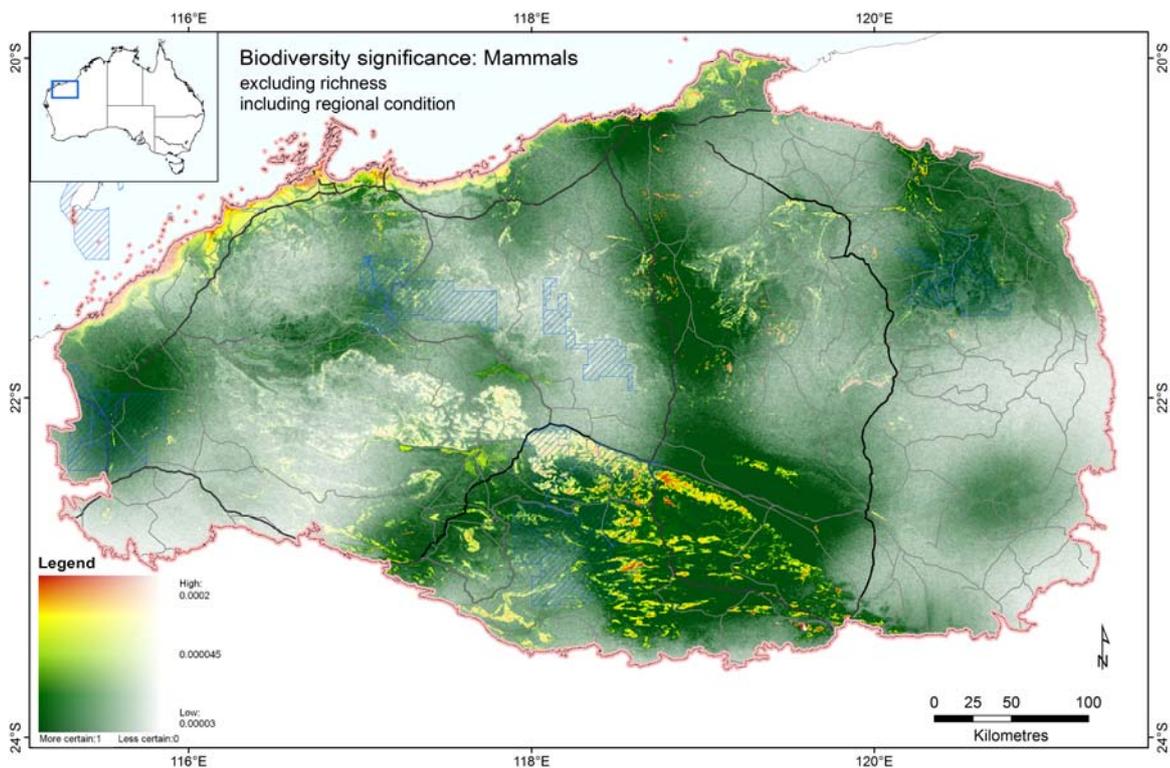
Figure 29 Biodiversity significance excluding richness and including regional condition for Mammals, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
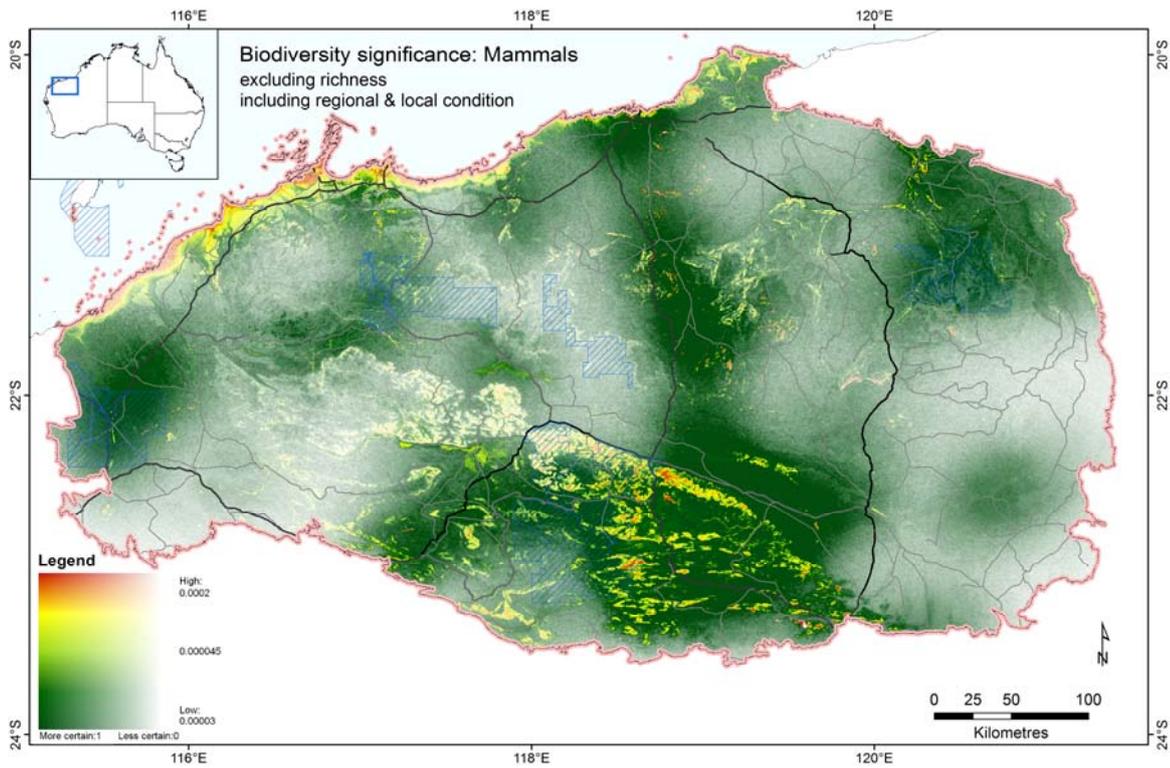
Figure 30 Biodiversity significance excluding richness and including regional and local condition for Mammals, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
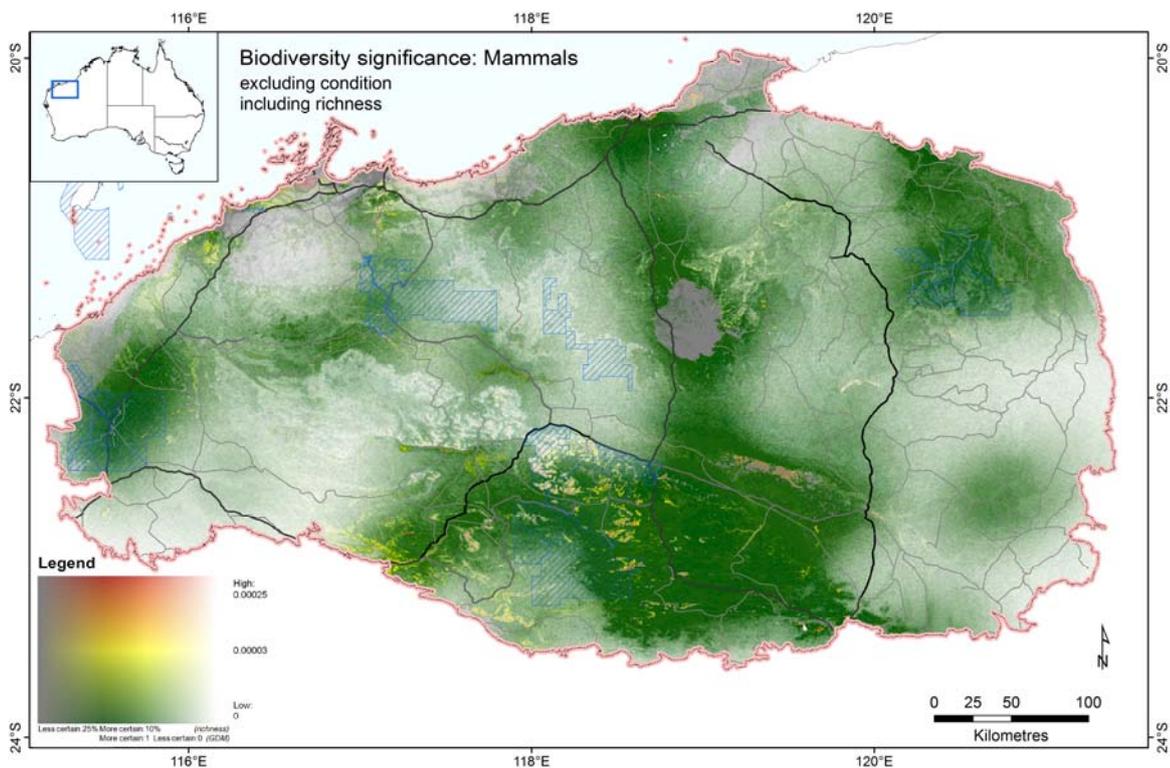
Figure 31 Biodiversity significance including richness excluding condition for all Mammals, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
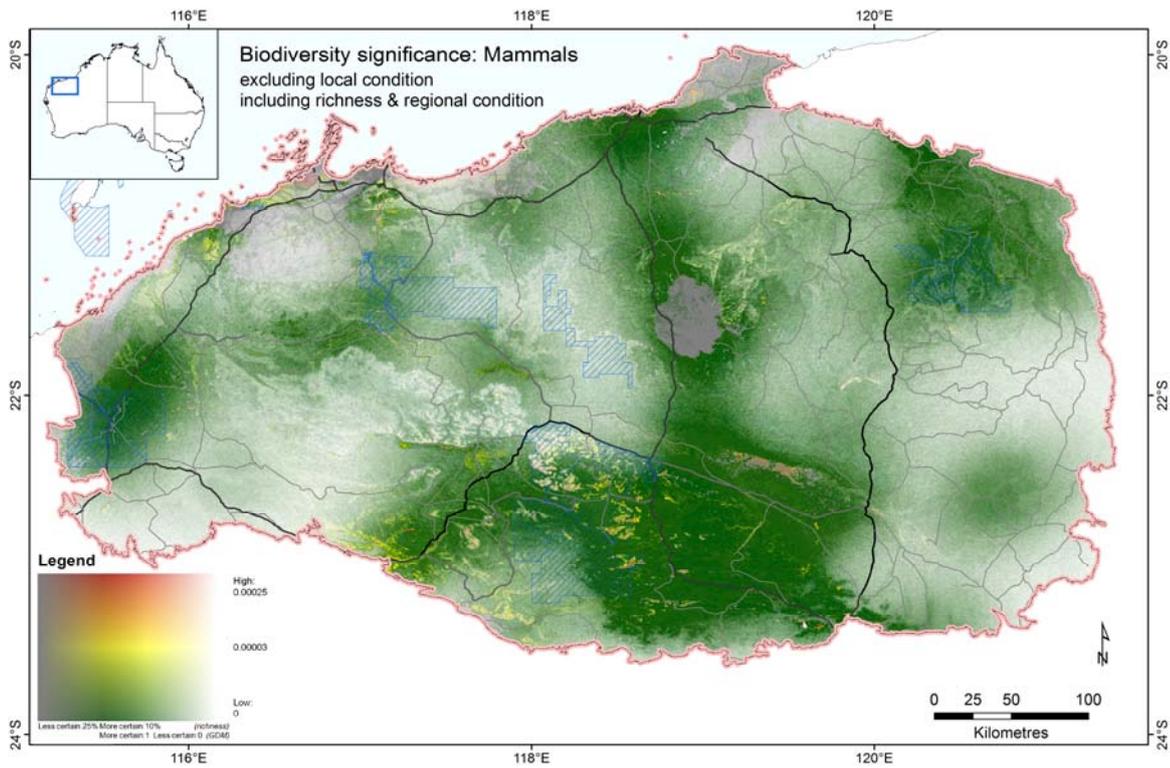
Figure 32 Biodiversity significance including richness and regional condition for Mammals, based on community-level modelling. Significance calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
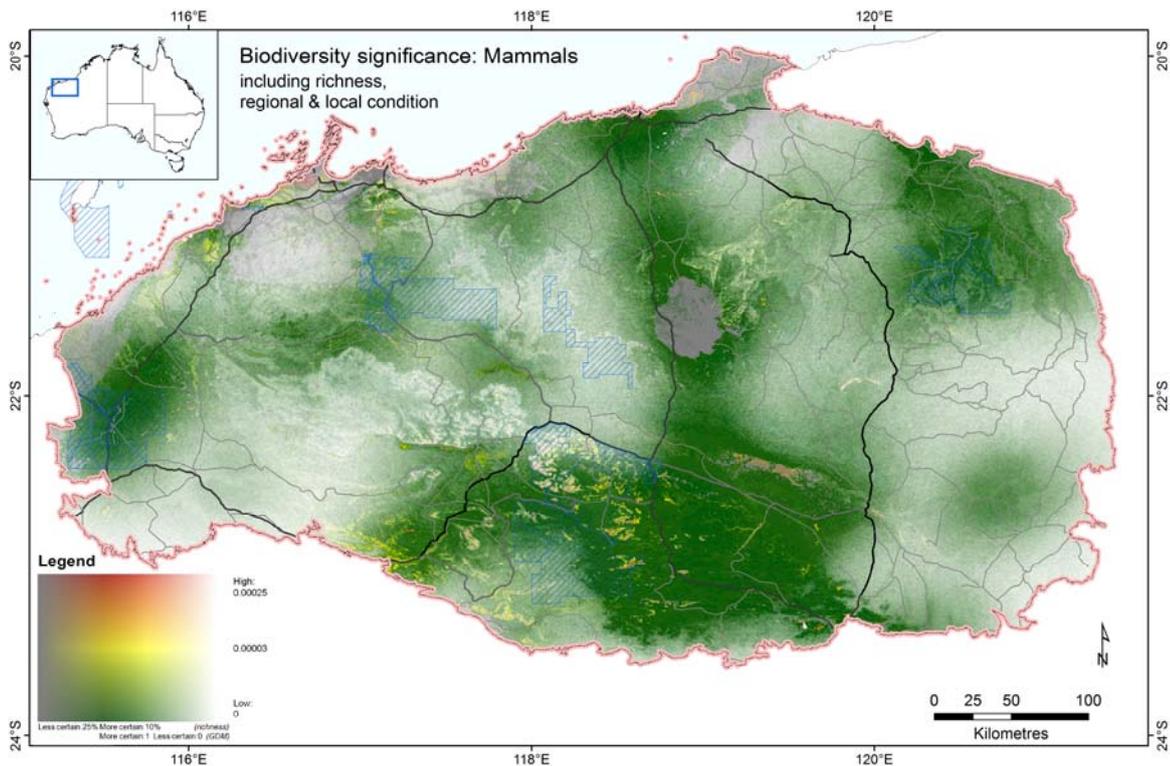
Figure 33 Biodiversity significance including richness, regional & local condition for Mammals, based on community-level modelling. Significance is here calculated as the scaled effect of removing each cell in degraded (c) condition from a degraded habitat. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
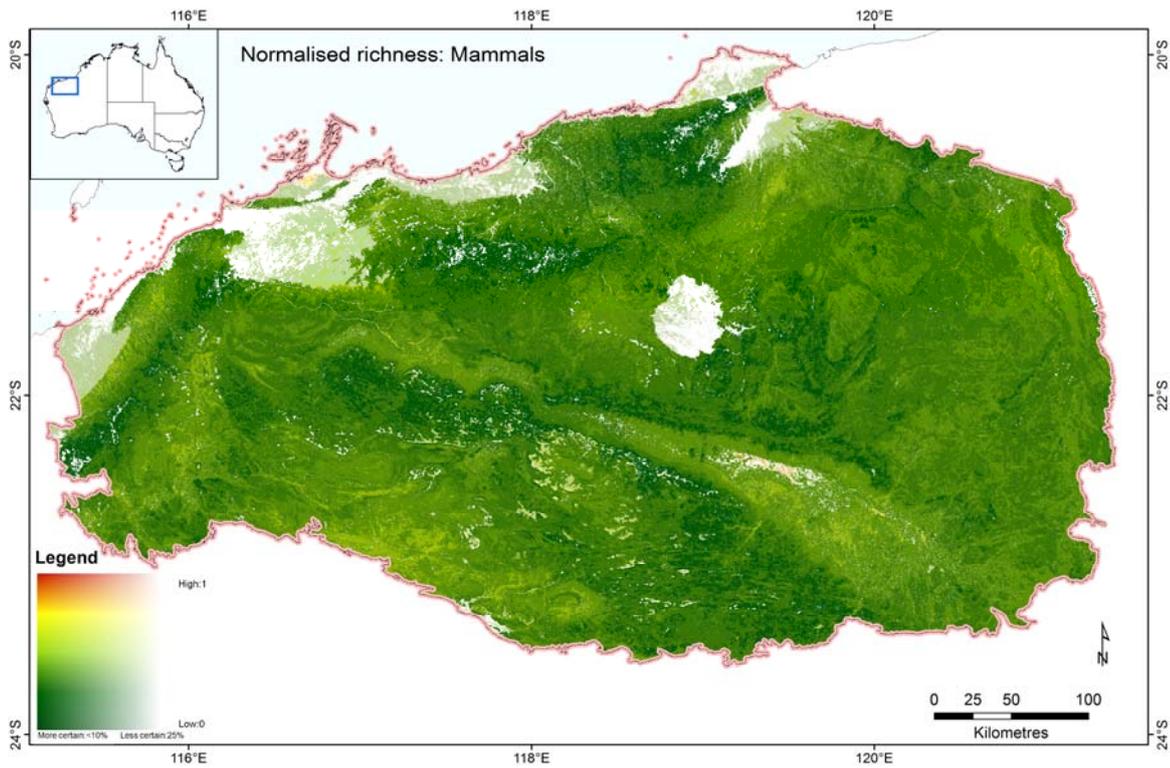
Figure 34 Normalised species richness (log fraction of maximum richness) for Mammals, based on community-level modelling. This is the right hand side, ln(r)/ln(rmax) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
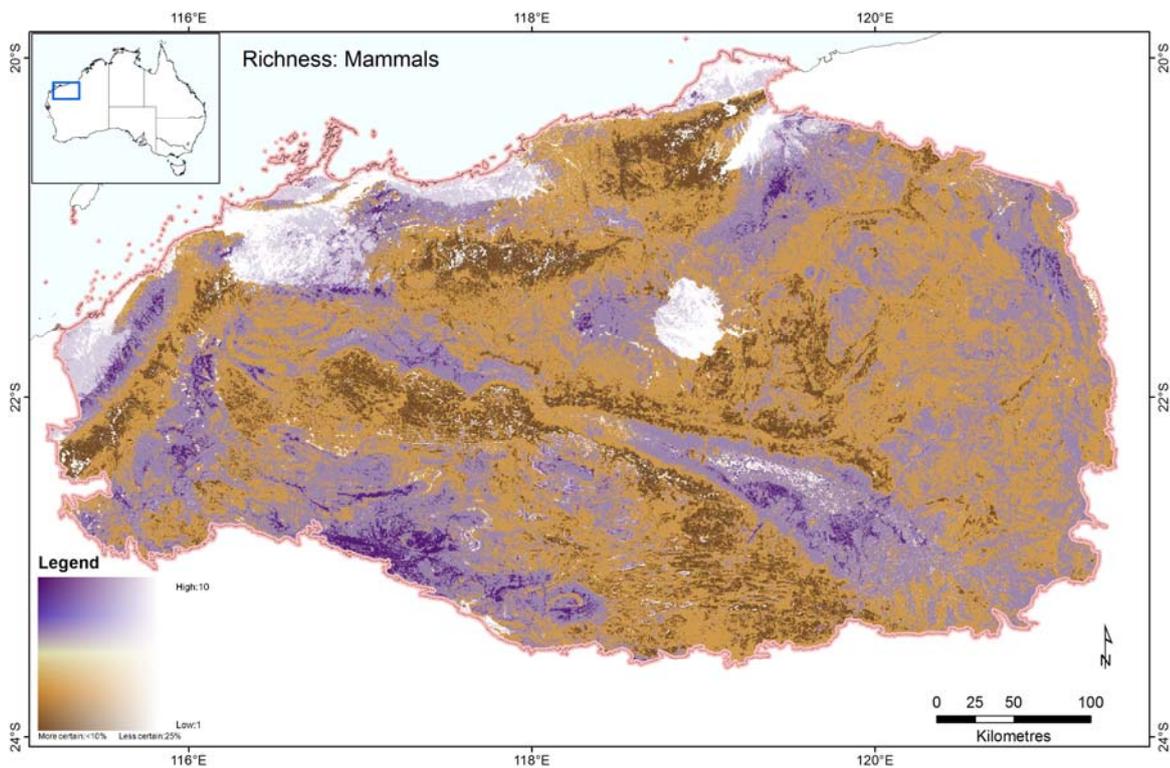
Figure 35 Absolute species richness (number of species in 9s grid cell) for Mammals, based on community-level modelling. Purple areas have more species than brown areas. Whiter areas are more uncertain than transparent areas.
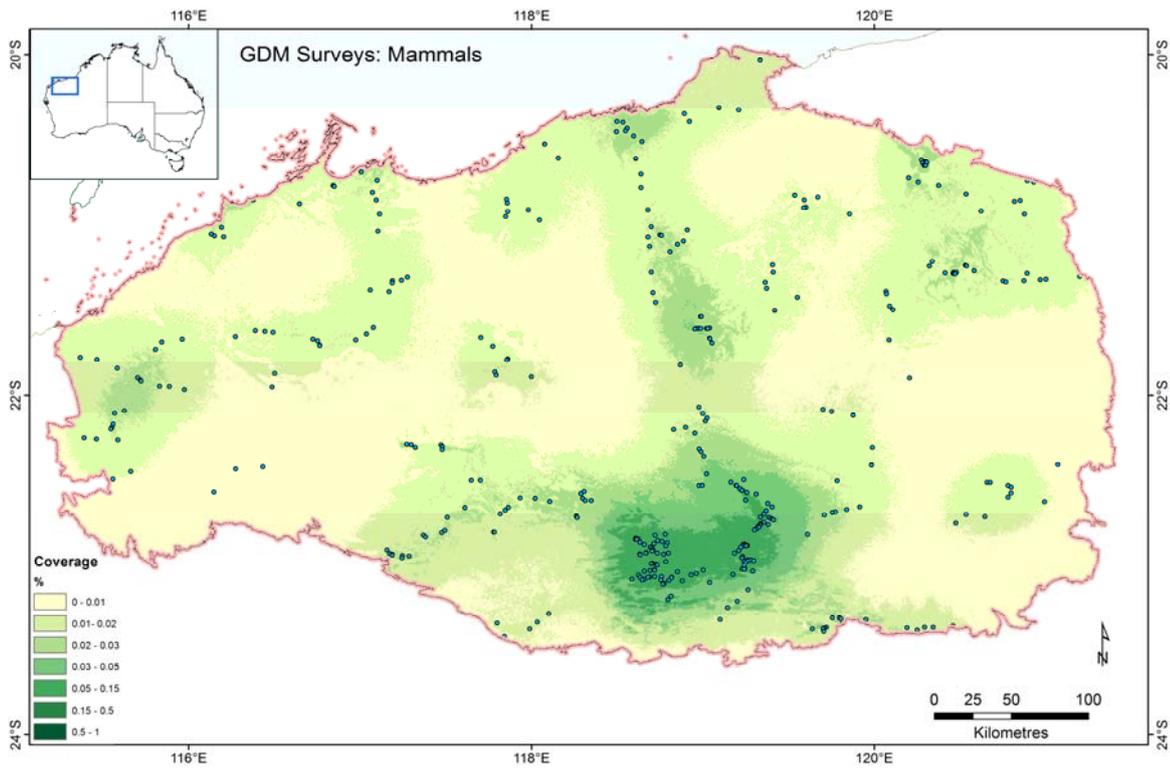
Figure 36 Location of surveys used in the Mammals GDM analysis (blue circles) overlaid on a map of sample density in GDM scaled environmental space. This describes the proportion of similar habitat for each cell i which is covered by survey sites, ranging from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the GDM uncertainty cloud applied in biodiversity significance maps.
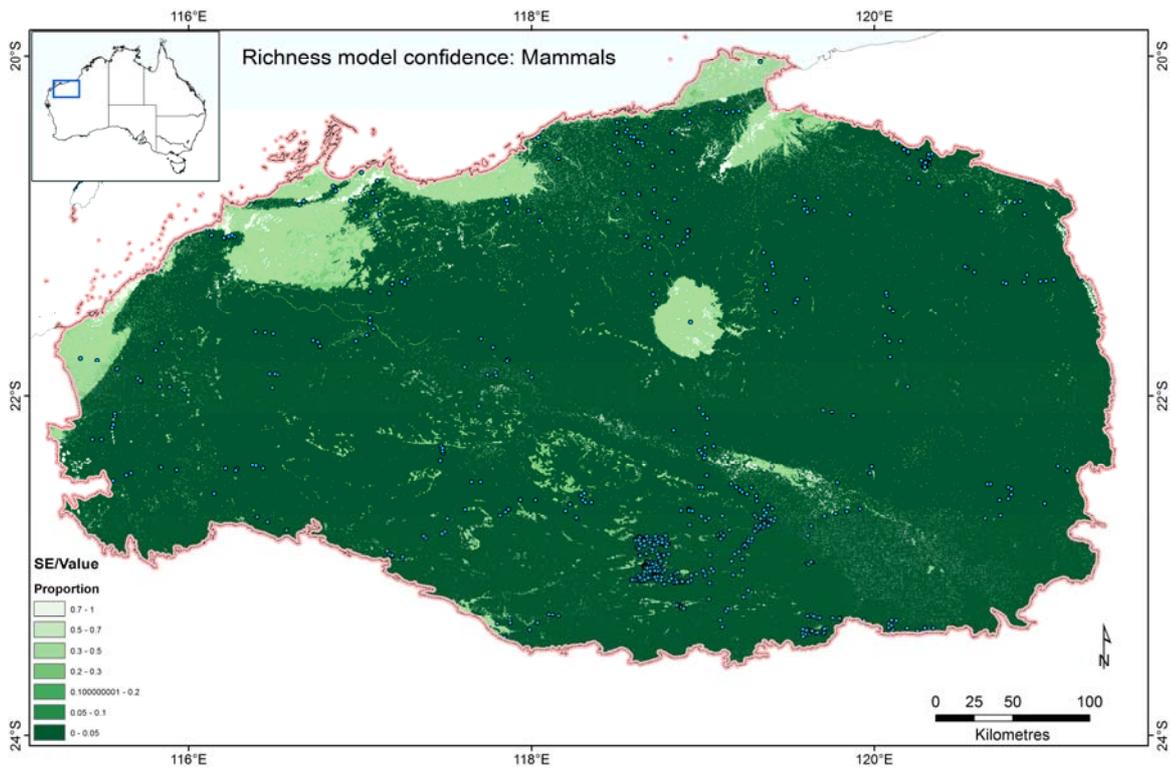
Figure 37 Location of surveys used in the Mammals richness analysis (blue circles) overlaid on a map of richness model standard error/value. Values range from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the richness model uncertainty cloud applied in biodiversity significance maps.

Vascular Plants
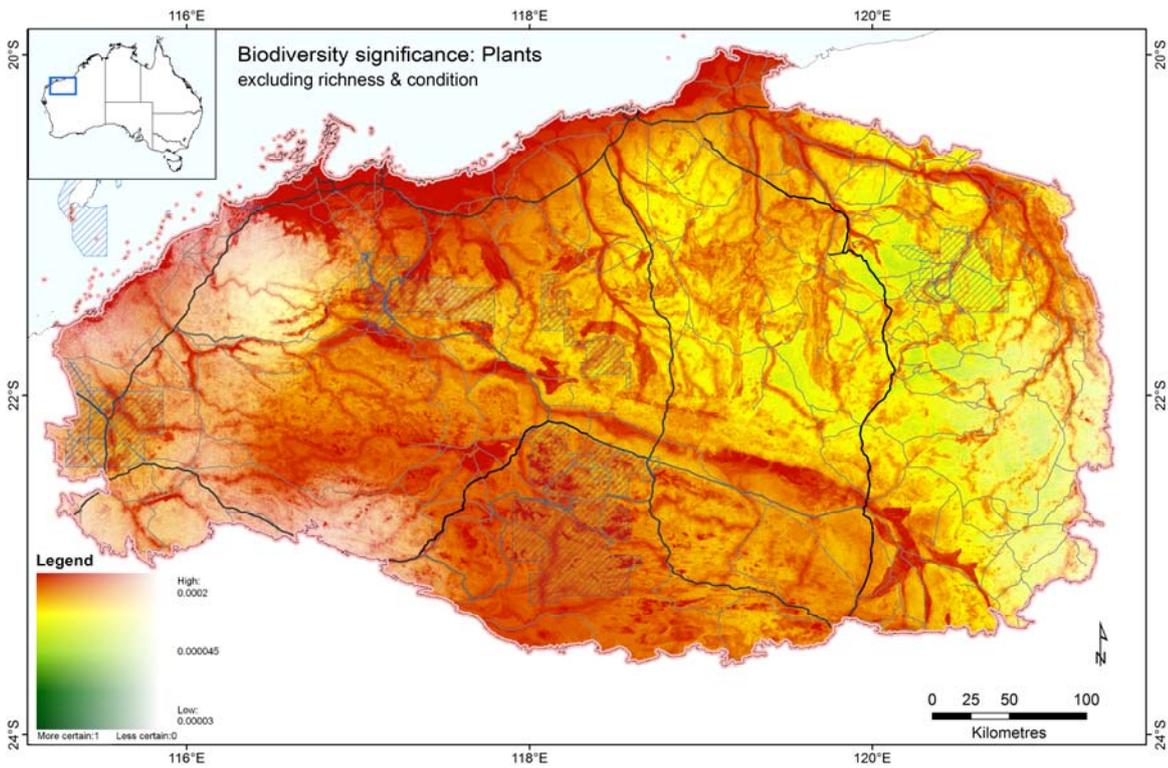
Figure 38 Biodiversity significance excluding richness and condition for Vascular Plants, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.

Figure 39 Biodiversity significance excluding richness and including regional condition for Vascular Plants, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.

Figure 40 Biodiversity significance excluding richness and including regional and local condition for Vascular Plants, based on community-level modelling. Significance calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.

Figure 41 Biodiversity significance including richness excluding condition for Vascular Plants, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.

Figure 42 Biodiversity significance including richness and regional condition for Plants, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
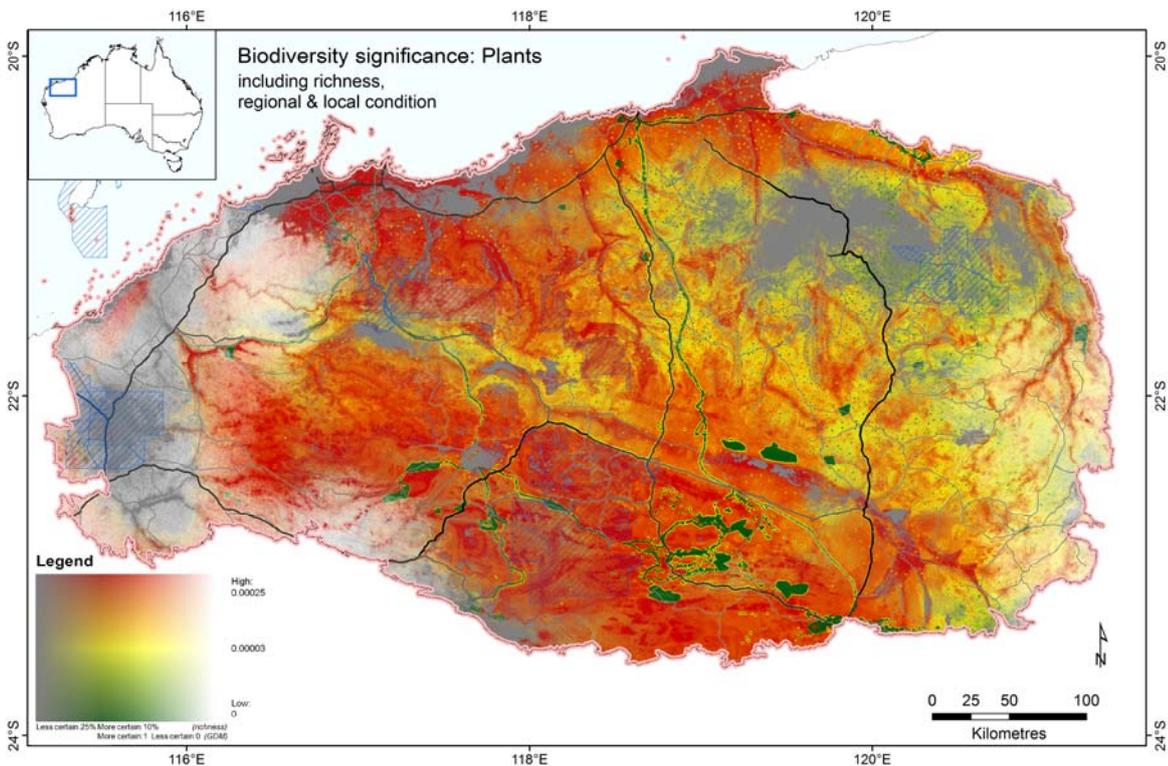
Figure 43 Biodiversity significance including richness, regional & local condition for Plants, based on community-level modelling. Significance is the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
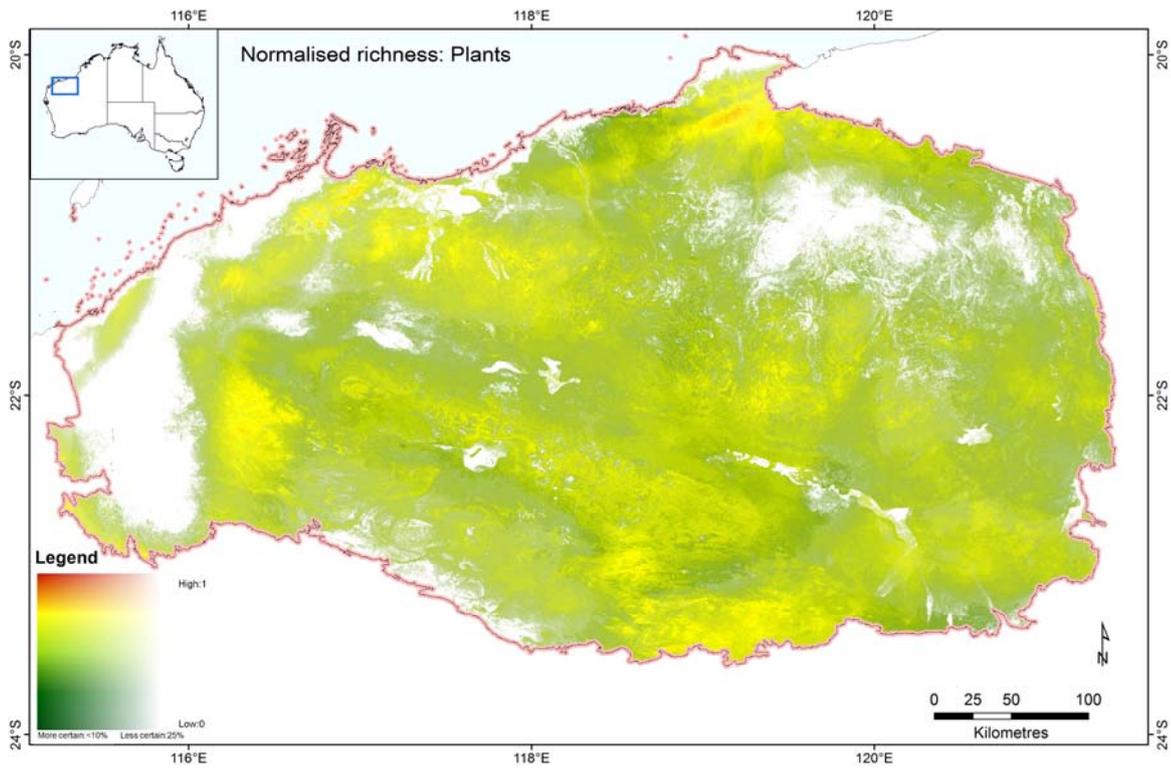
Figure 44 Normalised species richness (log fraction of maximum richness) for Vascular Plants, based on community-level modelling. This is the right hand side, ln(r)/ln(rmax) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
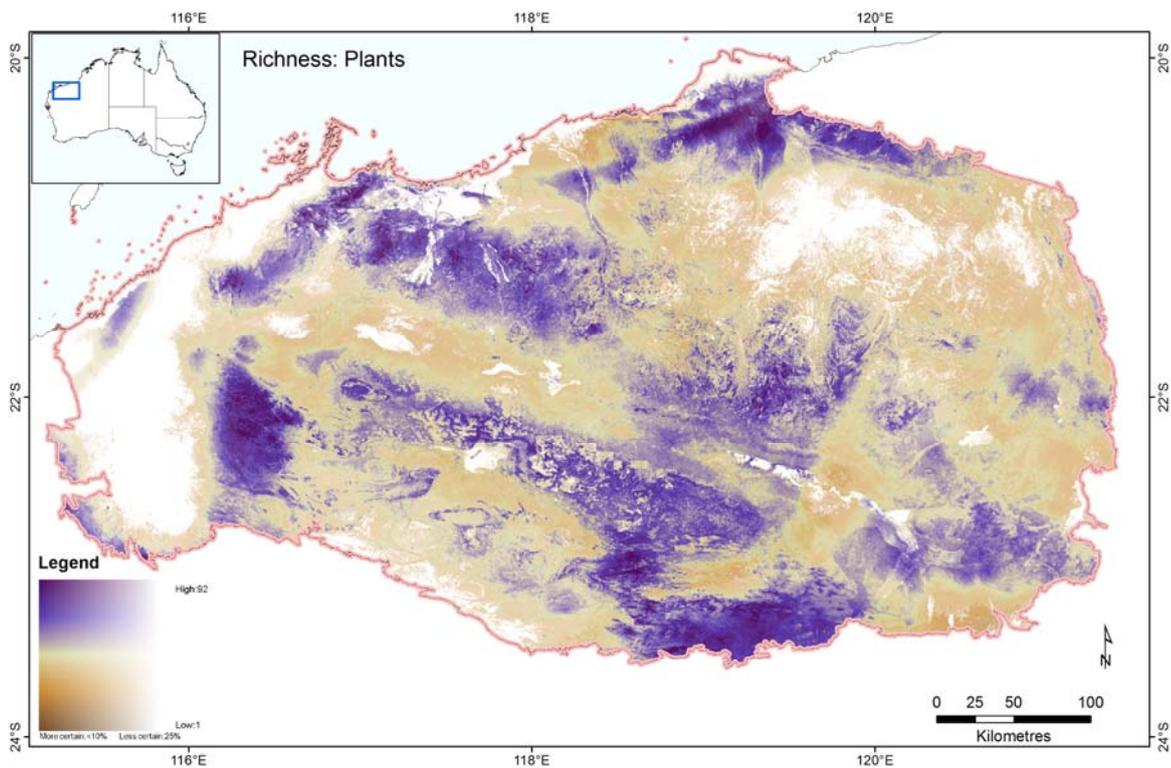
Figure 45 Absolute species richness (number of species in 9s grid cell) for Vascular Plants, based on community-level modelling. Purple areas have more species than brown areas. Whiter areas are more uncertain than transparent areas.
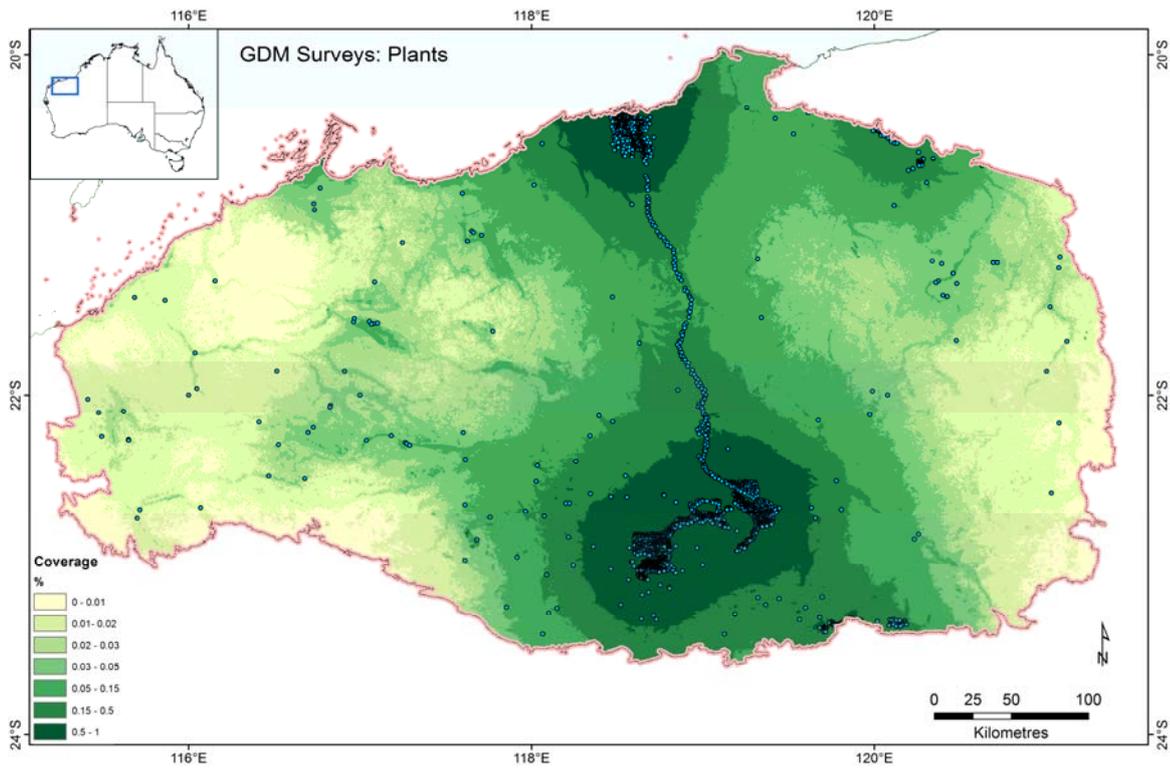
Figure 46 Location of surveys used in the Vascular Plants GDM analysis (blue circles) overlaid on a map of sample density in GDM scaled environmental space. This describes the proportion of similar habitat for each cell i which is covered by survey sites, ranging from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the GDM uncertainty cloud applied in biodiversity significance maps.
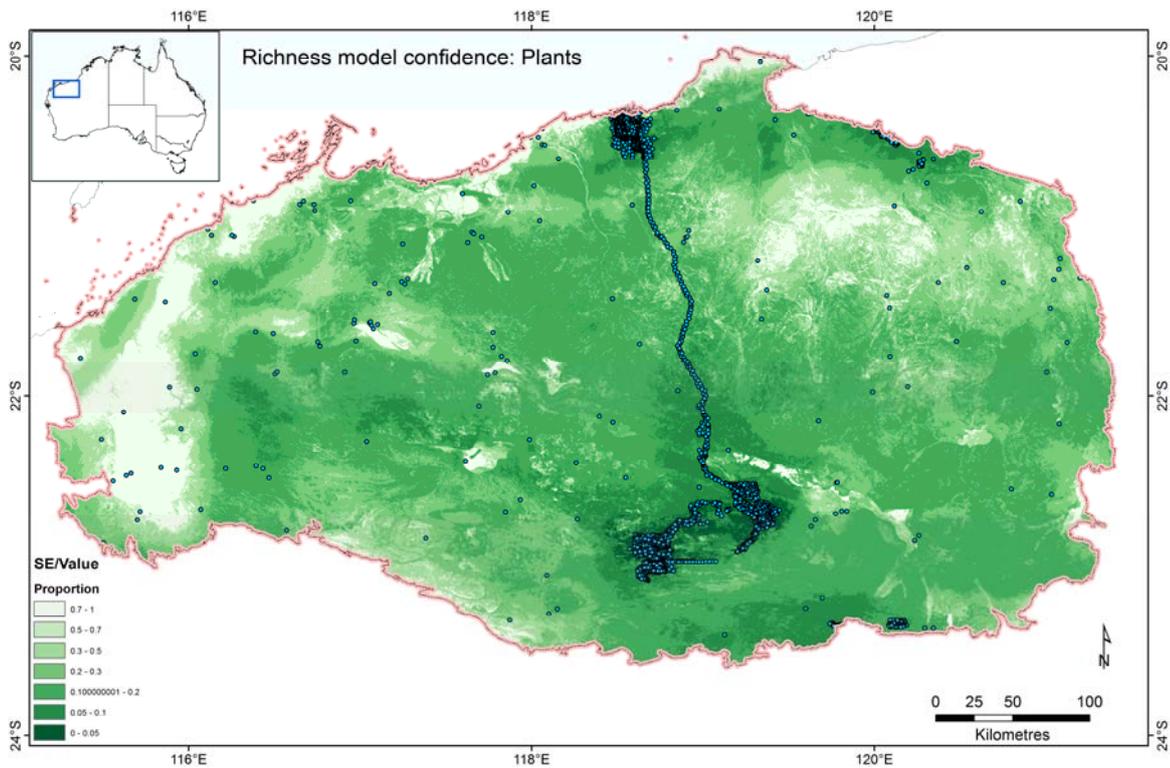
Figure 47 Location of surveys used in the Vascular Plants richness analysis (blue circles) overlaid on a map of richness model standard error/value. Values range from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the richness model uncertainty cloud applied in biodiversity significance maps.

Reptiles
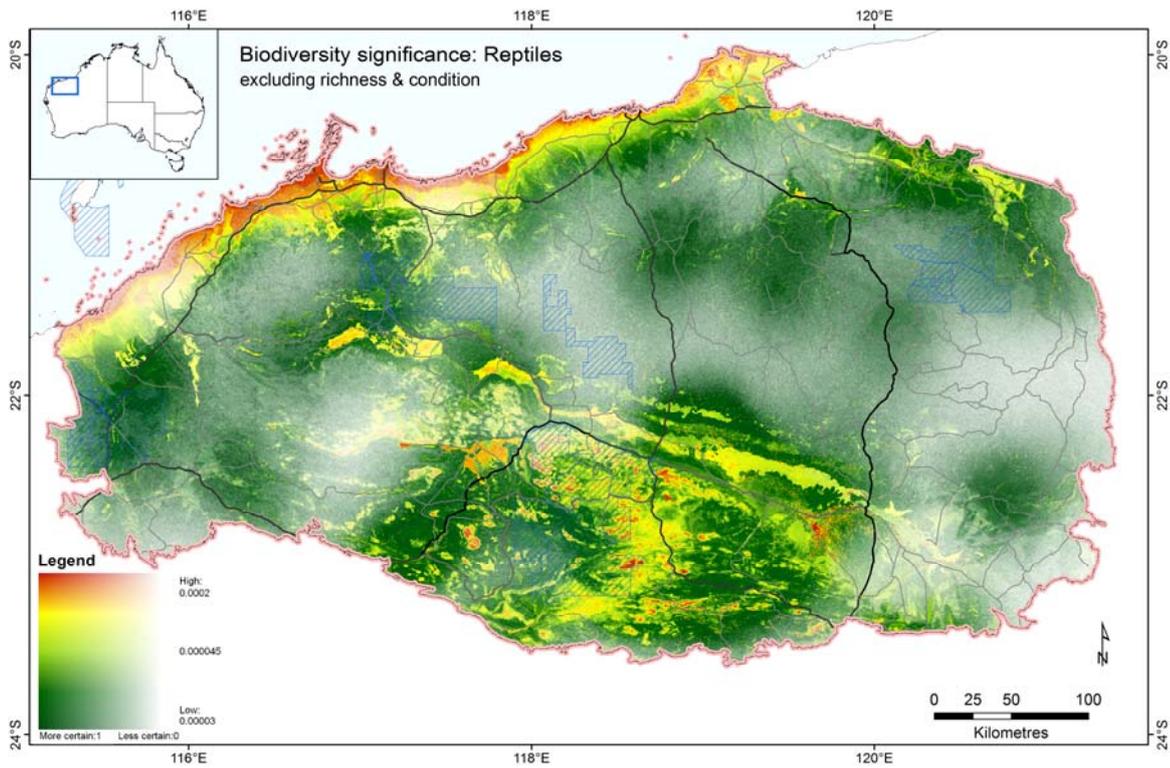
Figure 48 Biodiversity significance excluding richness and condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
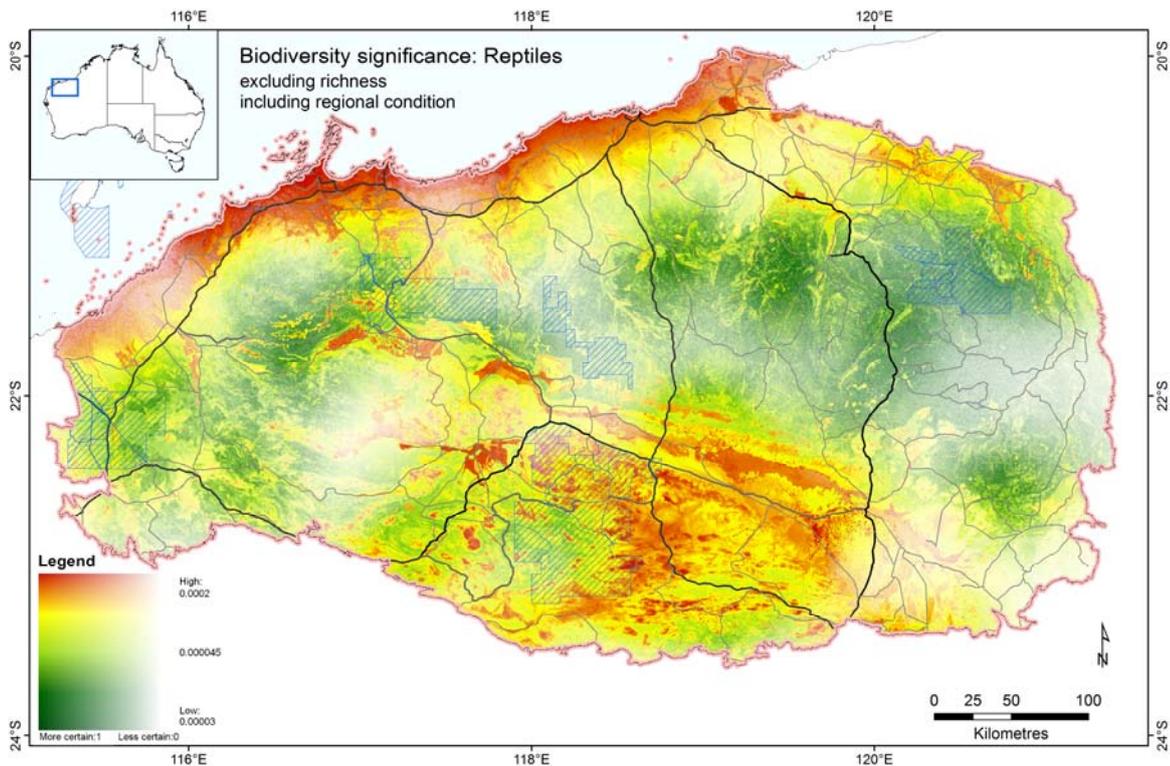
Figure 49 Biodiversity significance excluding richness and including regional condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
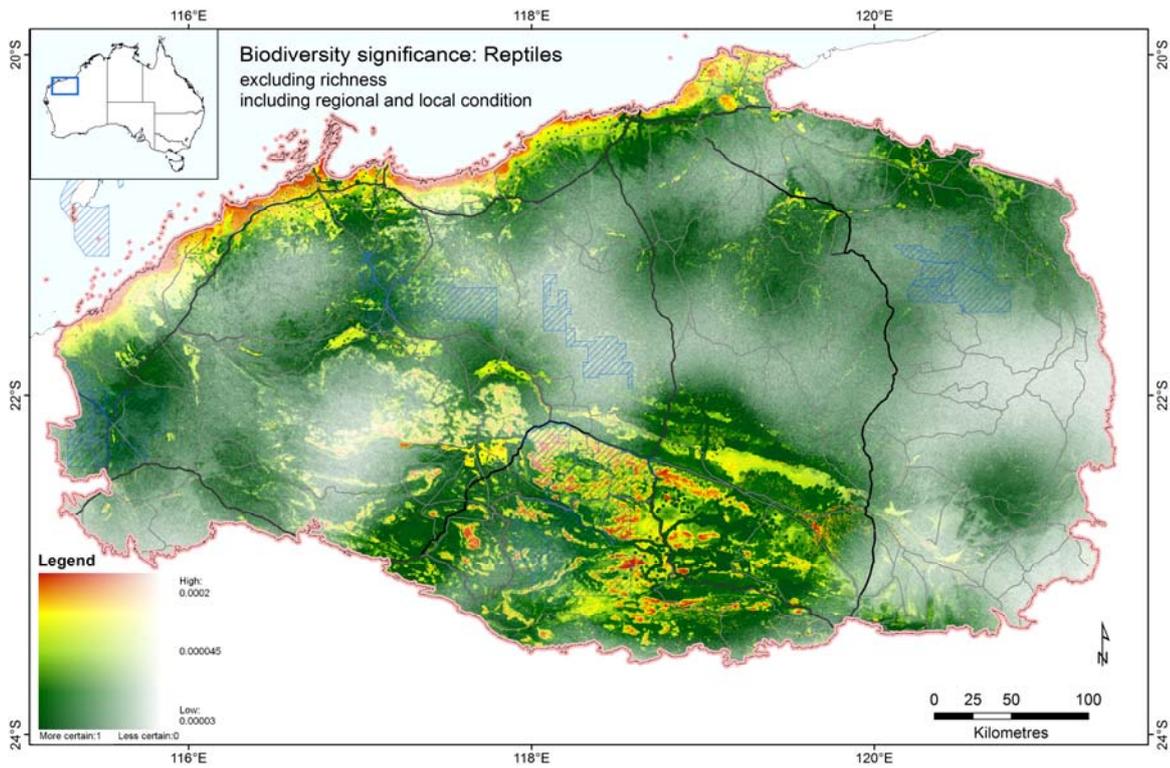
Figure 50 Biodiversity significance excluding richness and including regional and local condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (assuming local condition from interim layer) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
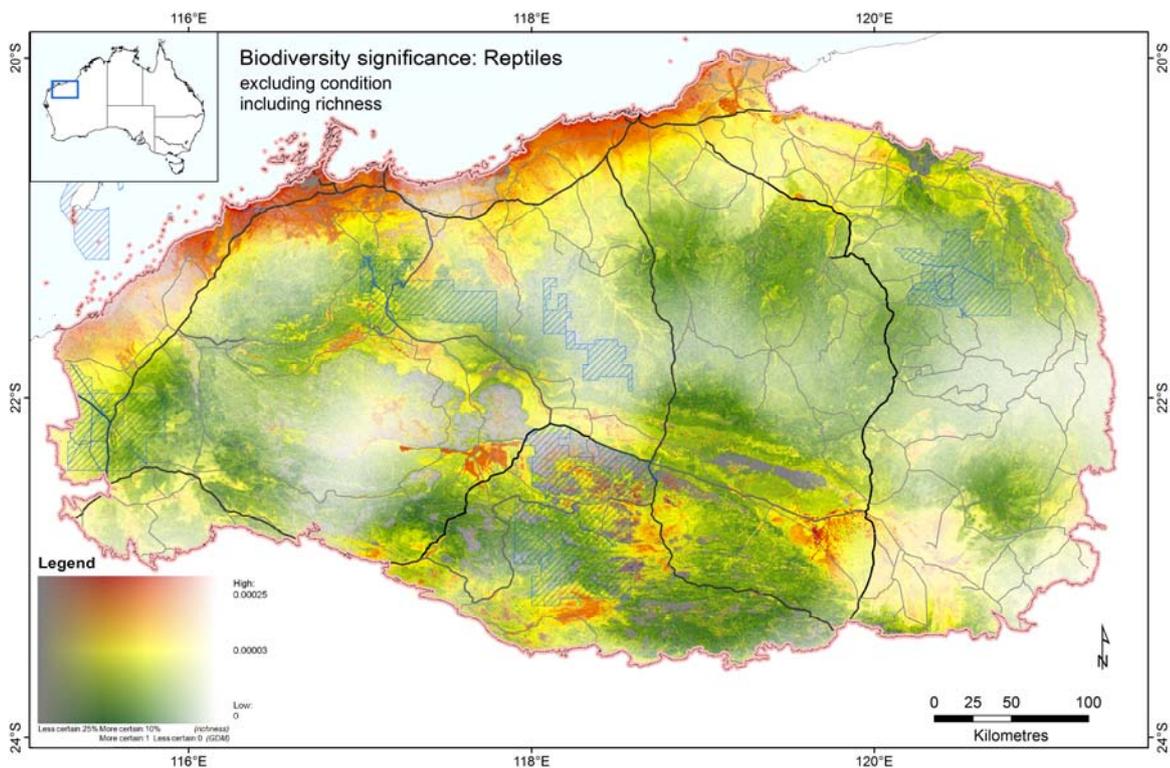
Figure 51 Biodiversity significance including richness excluding condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell as if the entire region were still in pristine condition. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
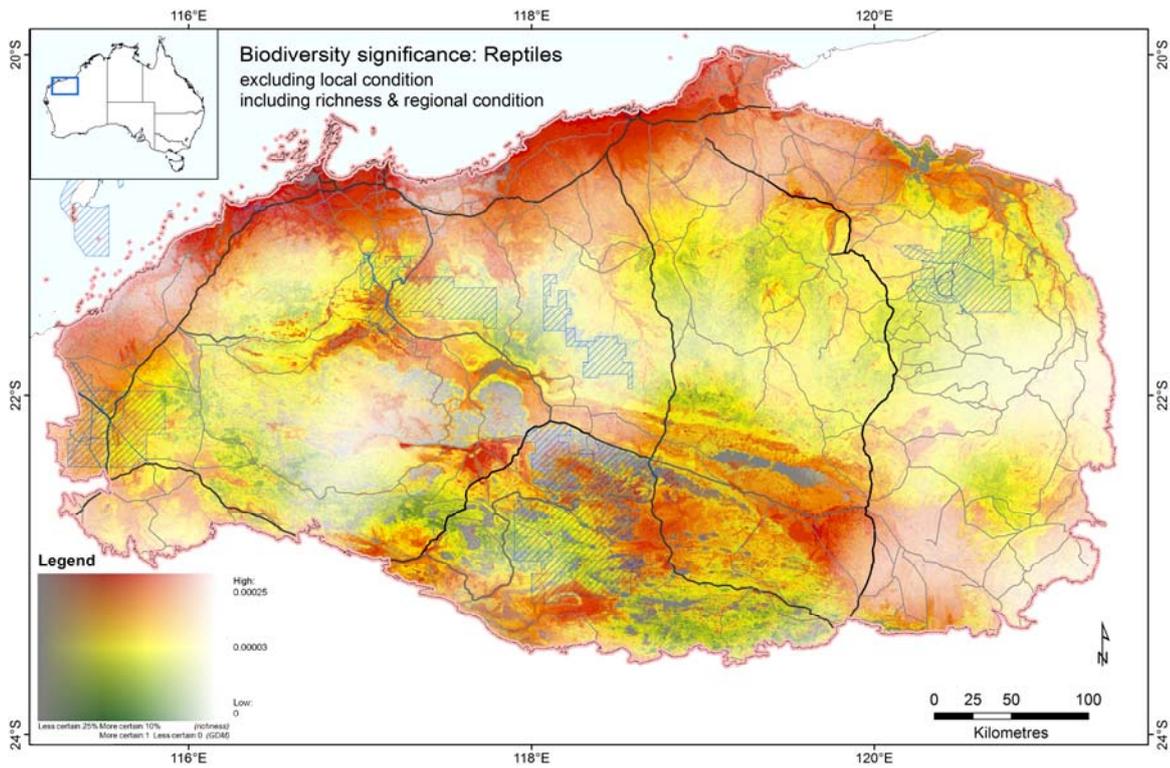
Figure 52 Biodiversity significance including richness and regional condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
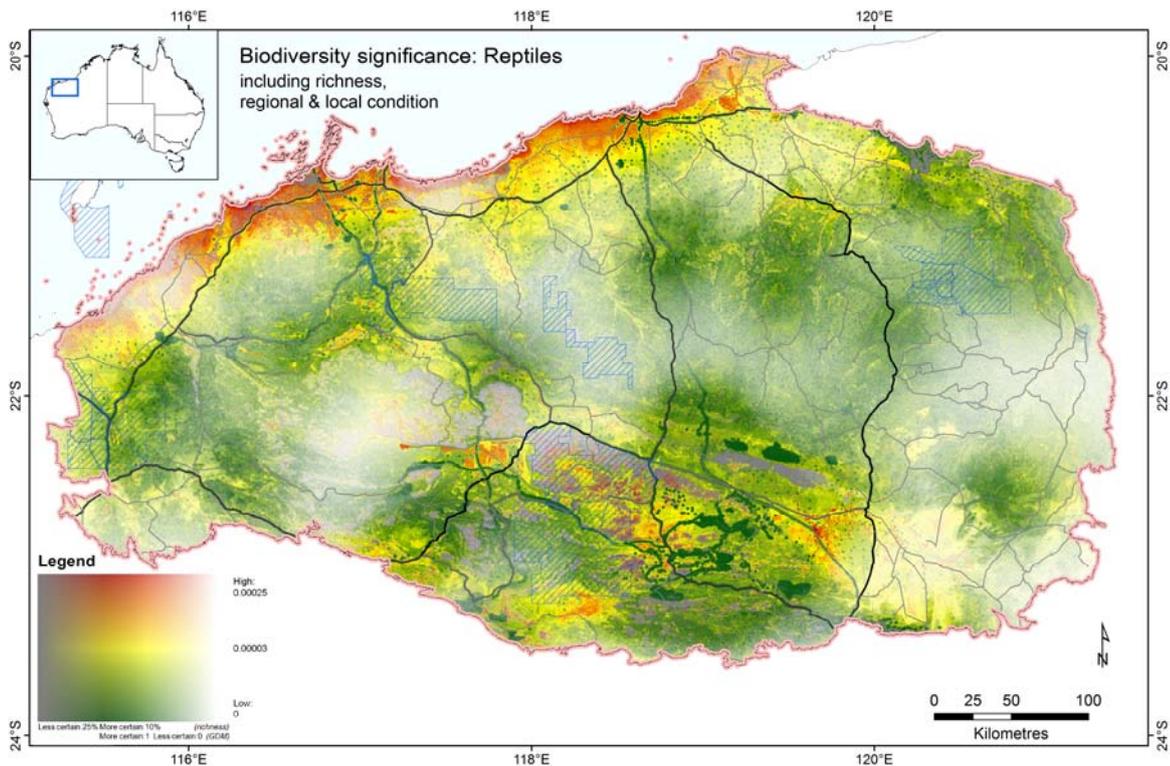
Figure 53 Biodiversity significance including richness and regional condition for Reptiles, based on community-level modelling. Significance is here calculated as the species-area scaled effect of removing each cell (as if local condition were still pristine) from the region in its present state. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter (GDM) or greyer (richness) areas are more uncertain than transparent areas.
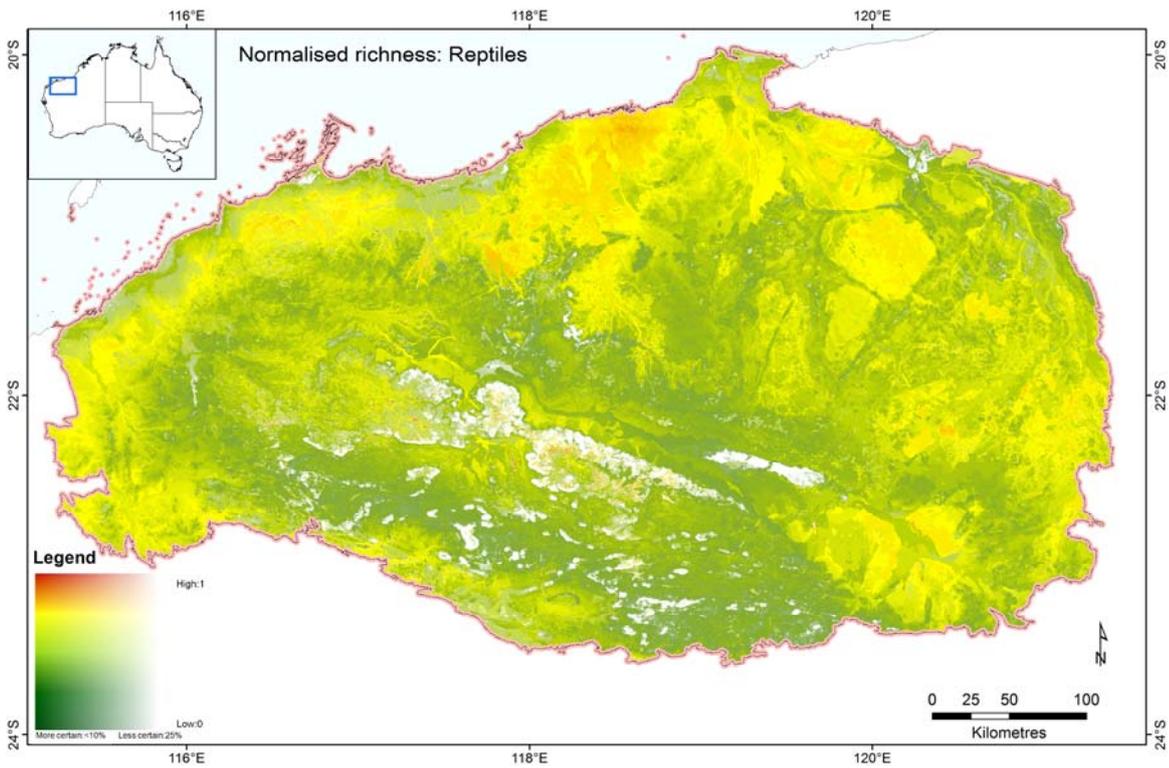
Figure 54 Normalised species richness (log fraction of maximum richness) for Reptiles, based on community-level modelling. This is the right hand side, ln(r)/ln(rmax) of the biodiversity significance equation. Darker green areas have a lower significance for biodiversity than yellow or red areas. Whiter areas are more uncertain than transparent areas.
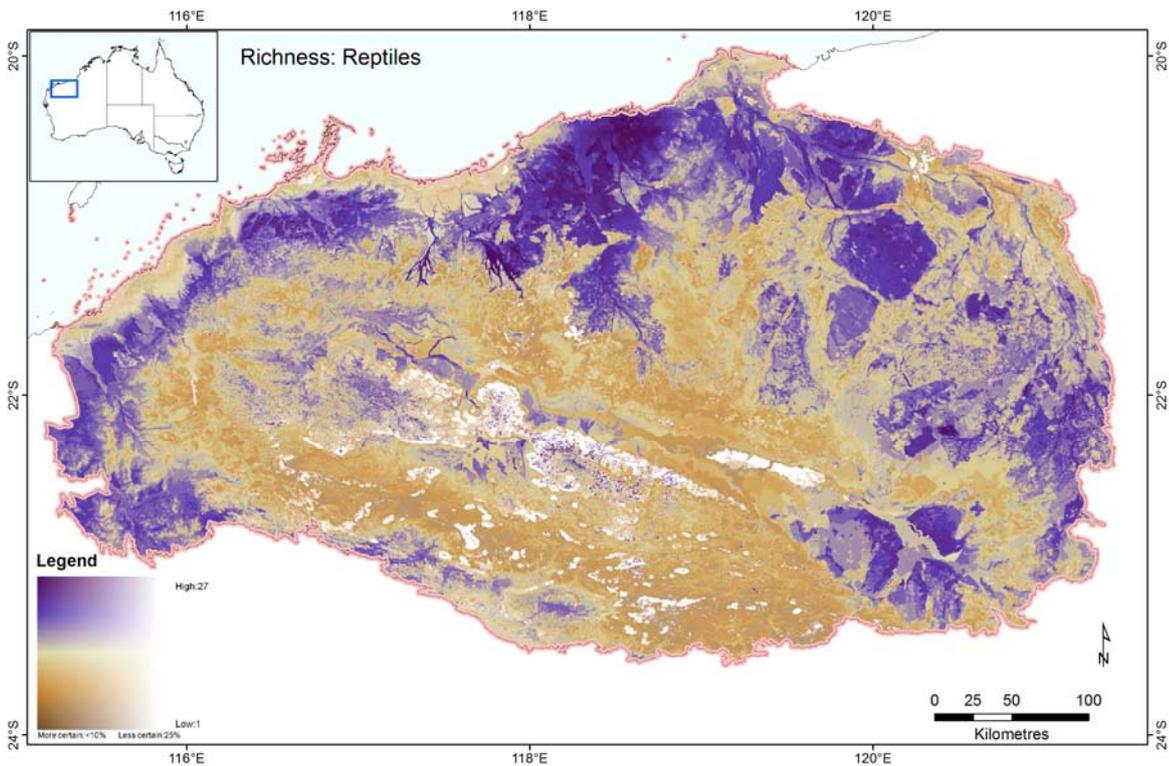
Figure 55 Absolute species richness (number of species in 9s grid cell) for Reptiles, based on community-level modelling. Purple areas have more species than brown areas. Whiter areas are more uncertain than transparent areas.
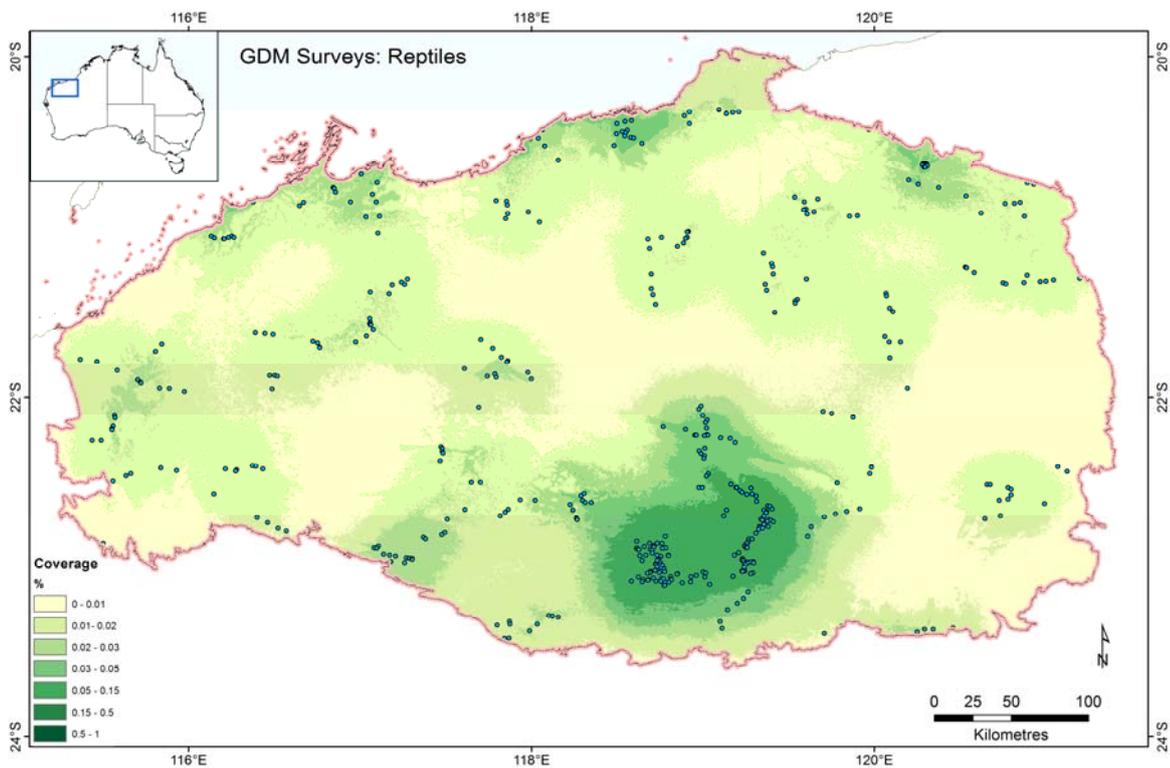
Figure 56 Location of surveys used in the Reptiles GDM analysis (blue circles) overlaid on a map of sample density in GDM scaled environmental space. This describes the proportion of similar habitat for each cell *i* which is covered by survey sites, ranging from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the GDM uncertainty cloud applied in biodiversity significance maps.
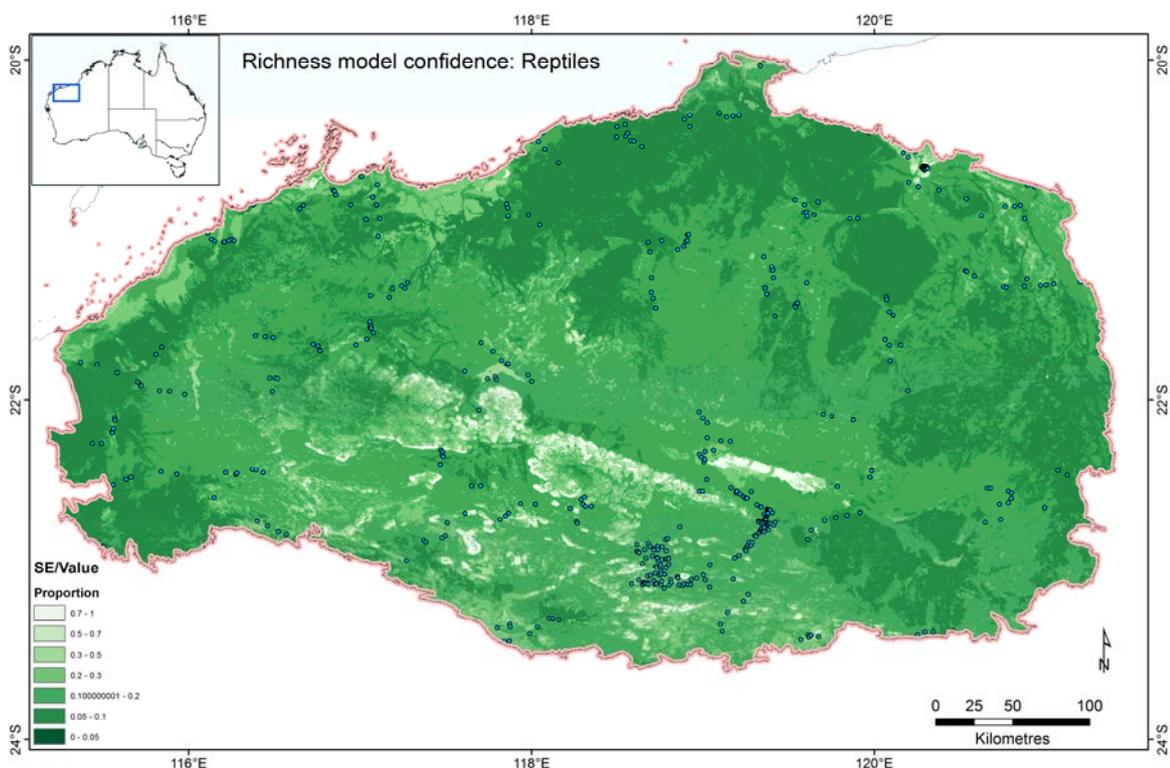
Figure 57 Location of surveys used in the Reptiles richness analysis (blue circles) overlaid on a map of richness model standard error/value. Values range from white/yellow (low coverage <0.001%) to dark green (good coverage >0.5%). This surface was used as the basis for the richness model uncertainty cloud applied in biodiversity significance maps.

References

Abbott P.F. & Tabony R.C. (1985). The estimation of humidity parameters. *Meteorological Magazine*, 114, 49-56.

Allen R., Pereira L., Raes D. & Smith M. (1998). Crop evapotranspiration - Guidelines for computing crop water requirements. In: *FAO Irrigation and Drainage Paper*. FAO - Food and Agriculture Organization of the United Nations Rome.

Allnutt T.F., Ferrier S., Manion G., Powell G.V.N., Ricketts T.H., Fisher B.L., Harper G.J., Irwin M.E., Kremen C., Labat J.-N., Lees D.C., Pearce T.A. & Rakotondrainibe F. (2008). A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. *Conservation Letters*, 1, 173-181.

Armstrong D.P. (2005). Integrating the Metapopulation and Habitat Paradigms for Understanding Broad-Scale Declines of Species. *Conserv. Biol.*, 19, 1402-1410.

AusCover (2013). AusCover data portal. URL http://data.auscover.org.au/xwiki/bin/view/Product+pages/Landsat+Fractional+Cover#HAbstractor Summary

Bailey D.W. (2005). Identification and Creation of Optimum Habitat Conditions for Livestock. *Rangeland Ecology & Management*, 58, 109-118.

Barry S.C. & Welsh A.H. (2002). Generalized additive modelling and zero inflated count data. *Ecol. Modell.*, 157, 179-188.

Bastin G., Pickup G., Chewings V. & Pearce G. (1993). Land degradation assessment in central Australia using a grazing gradient method. *Rangeland Journal*, 15, 190-216.

Bastin G., Scarth P., Chewings V., Sparrow A., Denham R., Schmidt M., O'Reagain P., Shepherd R. & Abbott B. (2012). Separating grazing and rainfall effects at regional scale using remote sensing imagery: A dynamic reference-cover method. *Remote Sens. Environ.*, 121, 443-457.

Beckett K.A. (2003). AIRBORNE GEOPHYSICS APPLIED TO GROUNDWATER MODELLING. In: *Advances in Regolith* (ed. I.C. R). Cooperative Research Centre for Landscape Environments and Mineral Exploration, pp. 8-10.

Beckett K.A. (2007). Multispectral analysis of high spatial resolution 256-channel radiometrics for soil and regolith mapping. In.

Belbin L., Marshall C. & Faith D. (1983). Representing relationships by automatic assignment of color. *Australian Computer Journal*, 15, 160-163.

Berry S., Mackey B. & Brown T. (2007). Potential applications of remotely sensed vegetation greenness to habitat analysis and the conservation of dispersive fauna. *Pacific Conservation Biology*, 13, 120-127.

Berry S.L. & Roderick M.L. (2006). Changing Australian vegetation from 1788 to 1988: effects of CO_2 and land-use change. *Australian Journal of Botany*, 54, 325-338.

Bowman D. (2012). Biodiversity crisis demands bolder thinking than bagging national parks. In: (ed. Conversation T) The Conversation

Breiman L. (2001). Random Forests. *Machine Learning*, 45, 5-32.

Breiman L., Cutler A., Liaw A. & Wiener M. (2013). Package 'randomForest': Breiman and Cutler's random forests for classification and regression. In. Cran-R Project online, http://cran.r-project.org/web/packages/randomForest/index.html.

Briggs I.C. (1974). Machine contouring using minimum curvature. *Geophysics*, 39, 39-48.

Burbidge A.H., Johnstone R.E. & Pearson D.J. (2009). Birds in a vast arid upland: avian biogeographical patterns in the Pilbara region of Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 247-270.

Bureau of Meteorology (2010). *Australian Hydrological Geospatial Fabric (Geofabric) Product Guide, Version 1.0*. Australian Government, Canberra.

Chilcott C.R., McCallum B.S., Quirk M.F. & Paton C.J. (2003). *Grazing Land Management Education Package Workshop Notes – Burdekin*. Meat & Livestock Australia Limited, Sydney.

Claridge J., Williams K.J. & Storey R.J.L. (2000). *Creation of the South-East Queensland depth index rescaled using CTI*. Enhanced Resource Assessment 2000-05. A JVAP project QDN3A Technical Report. Queensland Department of Natural Resources, Brisbane.

Clark D.A. (1997). Magnetic petrophysics and magnetic petrology: aids to geological interpretation of magnetic surveys. *AGSO Journal of Australian Geology and Geophysics*, 17, 83-103.

Crowley G.M. & Garnett S.T. (1998). Vegetation change in the grasslands and grassy woodlands of east-central Cape York Peninsula, Australia. *Pacific Conservation Biology*, 4, 132-148.

Crowley G.M. & Garnett S.T. (2001). Growth, seed production and effect of defoliation in an early flowering perennial grass, <i>Alloteropsis semialata</i> (Poaceae), on Cape York Peninsula, Australia. *Australian Journal of Botany*, 49, 735-743.

Czekanowski J. (1932). Coefficient of racial likeness, und durchschnitttliche differenz. *Anthropol Anz*, 9, 227-249.

Danaher T., Scarth P., Armston J., Collet L., Kitchen J. & Gillingham S. (2010). Ecosystem Function in Savannas: Measurement and Modelling at Landscape to Global Scales. In: *Section 3. Remote Sensing of Biophysical and Biochemical Characteristics in Savannas How different remote sensing technologies contribute to measurement and understanding of savannas.* (ed. Francis Ta). Taylor and Francis.

DCCEE (2012). Metadata: Forest Extent and Change (v8) Area-Corrected, Aggregated Products. In. Australian Government Department of Climate Change & Energy Efficiency (DCCEE) Canberra, p. 3.

De Vries R. (2009). *Australia's Substrate Fertility - Version 0.8.* Department of the Environment, Water, Heritage and the Arts, Canberra.

Department of Water (2011). Pilbara pool mapping dataset. In. Western Australian Department of Water Perth, WA, Australia.

Donohue R.J., Harwood T.D., Williams K.J., Ferrier S. & McVicar T.R. (2013). Estimating habitat condition using time series remote sensing and ecological survey data. In. Final Report to the CSIRO Earth Observation Informatics Transformational Capability Platform Canberra, p. 48.

Doughty P., Rolfe J.K., Burbidge A.H., Pearson D.J. & Kendrick P.G. (2009). Herpetological assemblages of the Pilbara biogeographic region, Western Australia: ecological associations, biogeographic patterns and conservation. *Records of the Western Australian Museum, Supplement*, 78, 315-341.

DSEWPAC (2012). Interim Biogeographic Regionalisation for Australia (IBRA), Version 7 (Bioregions and Subregions) In. Australian Government Department of Sustainability, Environment, Water, Population and Communities Canberra, Australia, p. 2.

Durrant B.J., Harvey M.S., Framena V.W., Ott R. & Waldock J.M. (2009). Patterns in the composition of ground-dwelling spider communities in the pilbara bioregion, Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 185-204.

Dyer R. (1997). Developing sustainable pasture management practices for the semi-arid tropics of the Northern Territory. In. Northern Territory Department of Primary Industry and Fisheries Katherine.

ESRI (2011). ArcGIS Desktop: Release 10. In. Environmental Systems Research Institute Redlands, CA, U.S.A.

Faith D.P., Ferrier S. & Williams K.J. (2008). Getting biodiversity intactness indices right: ensuring that "biodiversity" reflects "diversity". *Global Change Biology*, 14, 207-217.

Ferrier S. (2002). Mapping spatial pattern in biodiversity for regional conservation planning: Where to from here? *Systematic Biology*, 51, 331-363.

Ferrier S. & Drielsma M. (2010). Synthesis of pattern and process in biodiversity conservation assessment: a flexible whole-landscape modelling framework. *Diversity and Distributions*, 16, 386-402.

Ferrier S., Faith D.P., Arponen A. & Drielsma M. (2009). Community-level approaches to spatial conservation prioritization. In: *Spatial Conservation Prioritization - Quantitative Methods and Computational Tools* (eds. Moilanen A, Wilson KA & Possingham HP). Oxford University Press, pp. 94-109.

Ferrier S. & Guisan A. (2006). Spatial modelling of biodiversity at the community level. *Journal of Applied Ecology*, 43, 393-404.

Ferrier S., Manion G., Elith J. & Richardson K. (2007). Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. *Diversity and Distributions*, 13, 252-264.

Ferrier S., Powell G.V.N., Richardson K.S., Manion G., Overton J.M., Allnutt T.F., Cameron S.E., Mantle K., Burgess N.D., Faith D.P., Lamoreux J.F., Kier G., Hijmans R.J., Funk V.A., Cassis G.A., Fisher B.L., Flemons P., Lees D., Lovett J.C., Rompaey V. & R. R.S.A. (2004). Mapping More of Terrestrial Biodiversity for Global Conservation Assessment. *Bioscience*, 54, 1101-1109.

Fisher A., Hunt L.P., James C., Landsberg J., Phelps D.G., Smyth A. & Watson A. (2004). Review of total grazing pressure management issues and priorities for biodiversity conservation in rangelands: a resource to aid NRM planning. In: *Desert Knowledge CRC Project Report No. 3 (August 2004)*. Desert Knowledge CRC and Tropical Savannas CRC Alice Springs.

Fisher A. & Kutt A. (2007). Biodiversity and Land Condition in Tropical Savanna Rangelands: Technical Report. In. Tropical Savannas CRC, Darwin.

Flannery T. (2012). After the Future: Australia's New Extinction Crisis. *Quarterly Essay*.

Foran B.D., Bastin G.N., Remenga E. & Hyde K.W. (1982). The response to season, exclosure, and distance from water of three central Australian pasture types grazed by cattle. *Australian Rangeland Journal*, 4, 5-15.

Ford H.A., Barrett G.W., Saunders D.A. & Recher H.F. (2001). Why have birds in the woodlands of Southern Australia declined? *Biol. Conserv.*, 97, 71-88.

Franklin D.C. (1999). Evidence of disarray amongst granivorous bird assemblages in the savannas of northern Australia, a region of sparse human settlement. *Biol. Conserv.*, 90, 53-63.

Funk V.A., Richardson K.S. & Ferrier S. (2005). Survey-gap analysis in expeditionary research: where do we go from here? *Biological Journal of the Linnean Society*, 85, 549-567.

Furby S.L., Caccetta P.A., Wallace J.F., Lehmann E.A. & Zdunic K. (2009). Recent development in vegetation monitoring products from Australia's national carbon accounting system. In: *Igarss: 2009 IEEE International Geoscience and Remote Sensing Symposium - Earth Observation Origins to Application, 12-17 July 2009*. IEEE Xplore Cape Town, South Africa, pp. 276-279.

Gallant J. (2011). Building the national one second digital elevation model of Australia. In: *Water Information Research and Development Alliance (WIRADA) Science Symposium 1–5 August 2011*. CSIRO Water for a Healthy Country Melbourne, Australia.

Gallant J. & Austin J. (2012). Metadata: Topographic position index derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011a). Metadata: 300 m elevation range derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011b). Metadata: 300 m focal median of percent slope derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011c). Metadata: 1000 m elevation range derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011d). Metadata: Aspect derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011e). Metadata: Contributing Area - Multiple Flow Direction (Partial) derived from 1 second DEM-H. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011f). Metadata: Percent slope derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011g). Metadata: Plan curvature derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011h). Metadata: Profile curvature derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011i). Metadata: Slope relief derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J., Austin J. & Dowling T. (2011j). Metadata: Topographic Wetness Index (TWI) derived from 1 second DEM-S. In. CSIRO Land and Water Canberra, p. 3.

Gallant J. & Read A. (2009). Enhancing the SRTM data for Australia. In: *Proceedings of Geomorphometry, 31 August - 2 September 2009* (eds. Purves R, Gruber S, Straumann R & Hengl T). University of Zurich Zurich, pp. 149-154.

Gallant J.C. & Dowling T.I. (2003). A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resources Research*, 39, 1347.

George A.S., McKenzie N.L. & Doughty P. (2011). A biodiversity survey of the Pilbara region of Western Australia. *Records of the Western Australian Museum, Supplement 78 (Part A)*, 1-311.

Geoscience Australia (2009). *Gravity grid of Australia and surrounding areas (National geoscience dataset)*. Geoscience Australia, Canberra.

Geoscience Australia (2010). The Radiometric Map of Australia Dataset. In. Geoscience Australia, Australian Government Canberra, p. 2.

Geoscience Australia (2013a). Australian mines atlas. URL http://www.australianminesatlas.gov.au/?site=atlas

Geoscience Australia (2013b). Metadata: 3 Second SRTM Derived Smoothed Digital Elevation Model (DEM-S) Version 1.0. In. Geoscience Australia, Australian Government Canberra.

Geoscience Australia & CSIRO Land & Water (2010). 1 Second SRTM Derived Digital Elevation Models User Guide. Version 1.0. In: *1 second SRTM Level 2 Derived Smoothed Digital Elevation Model (DEM-S) Version 1.0*. Geoscience Australia, Australian Government Canberra.

Geoscience Australia & CSIRO Land & Water (2011). 1 second SRTM Derived Digital Elevation Models User Guide: 1 second DSM, DEM, DEM-S & DEM-H, 3 second DSM, DEM & DEM-S, Version 1.0.4. In. Geoscience Australia, Canberra.

Gibson L.A. & McKenzie N.L. (2009). Environmental associations of small ground-dwelling mammals in the Pilbara region, Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 91-122.

Google Inc. (2013). Google Earth (Version 7.1.1.1871). In.

Gray J., Bishop T., Smith P., Robinson N. & Brough D. (2012). A pragmatic quantitative relationship for soil organic carbon distribution in eastern Australia. In: *The 5th Global workshop on Digital Soil Mapping 2012 - Digital soil assessments and beyond, 10-13 April* Sydney, Australia.

Groves R.H., Hosking J.R., Batianoff G.N., Cooke D.A., Cowie I.D., Johnson R.W., Keighery G.J., Lepschi B.J., Mitchell A.A., Moerkerk M., Randall R.P., Rozefelds A.C., N.G. W. & Waterhouse B.M. (2003). *Weed categories for natural and agricultural ecosystem management*. Bureau of Rural Sciences, Australian Government, Canberra.

Guerschman J.P., Van Dijk A.I.J.M., Mattersdorf G., Beringer J., Hutley L.B., Leuning R., Pipunic R.C. & Sherman B.S. (2009). Scaling of potential evapotranspiration with MODIS data reproduces flux observations and catchment water balance observations across Australia. *J. Hydrol.*, 369, 107-119.

Gunn P.J. (1997). Regional magnetic and gravity responses of extensional sedimentary basins. *AGSO Journal of Australian Geology and Geophysics*, 17.

Guthrie N.A., Weir T. & Will K. (2009). Localised and regional patterns in ground-dwelling beetle assemblages in a semi-tropical arid zone environment. *Records of the Western Australian Museum, Supplement*, 78, 169-184.

Hall L.S., Krausman P.R. & Morrison M.L. (1997). The habitat concept and a plea for standard terminology. *Wildlife Soceity Bulletin*, 25, 173-182.

Harwood T., Ferrier S., Williams K.J., Liu J., Perry J. & Perkins G. (2013). Biodiversity significance analyses and associated uncertainty for five biological groups across the Pilbara provided as an ESRI Map Package. v1. Data Collection. In. CSIRO Data Access Portal https://www.data.csiro.au/dap/landingpage?pid=csiro:7591.

Hastie T.J. & Tibshirani R.J. (1990). *Generalized Additive Models*. Chapman & Hall/CRC.

Heterick B.E., Durrant B. & Gunawardene N.R. (2009). The ant fauna of the Pilbara Bioregion, Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 157-167.

Holechek J.L. (1988). An Approach for setting the stocking rate. *Rangelands*, 10-14, 10-.

Huey R.B., Deutsch C.A., Tewksbury J.J., Vitt L.J., Hertz P.E., Perez H.J.A. & Garland T. (2009). Why tropical forest lizards are vulnerable to climate warming. *Proceedings of the Royal Society B-Biological Sciences*, 276, 1939-1948.

Hutchinson M., Stein J., Stein J., Anderson H. & Tickle P. (2008). *GEODATA 9 second DEM and D8. Digital elevation model version 3 and flow direction grid. Gridded elevation and drainage data. Source scale 1:250 000. User guide (3rd ed)*. 3 edn. Fenner School of Environment and Society, the Australian National University and Geoscience Australia, Australian Government, Canberra.

Jacquier D. (2011a). Metadata: ASRIS 0-1m Plant Available Water Capacity (250m raster). In. CSIRO Land and Water Canberra.

Jacquier D. (2011b). Metadata: ASRIS 0-30cm Bulk Density (250m raster). In. CSIRO LAnd and Water Canberra.

Jacquier D. (2011c). Metadata: ASRIS 0-30cm Clay Content (250m raster). In. CSIRO Land and Water Canberra.

Jacquier D. (2011d). Metadata: ASRIS Australian Soil Classification - Dominant Soil Order (250m raster). In. CSIRO LAnd and Water Canberra.

James C.D. (2003). Response of vertebrates to fenceline contrasts in grazing intensity in semi-arid woodlands of eastern Australia. *Austral. Ecol.*, 28, 137-151.

Johnson C.N. (2006). *Australia's Mammal Extinctions. A 50 000 year History*. Cambridge University Press, Port Melbourne.

Kennard M.J., Pusey B.J., Olden J.D., MacKay S.J., Stein J.L. & Marsh N. (2010). Classification of natural flow regimes in Australia to support environmental flow management. *Freshw. Biol.*, 55, 171-193.

Kutt A.S. & Fisher A. (2011). Increased grazing and dominance of an exotic pasture (*Bothriochloa pertusa*) affects vertebrate fauna species composition, abundance and habitat in savanna woodland. *The Rangeland Journal*, 33, 49-58.

Kutt A.S. & Gordon I.J. (2012). Variation in terrestrial mammal abundance on pastoral and conservation land tenures in north-eastern Australian tropical savannas. *Animal Conservation*, 15, 416-425.

Kutt A.S. & Kemp J.E. (2011). Exotic pasture cover, climate and grazing pressure reduce native plant diversity in a tropical savanna. *The Rangelands Journal*, in review.

Kutt A.S., Vanderduys E.P., Perry J.J., Perkins G.C., Kemp J.E., Bateman B.L., Kanowski J. & Jensen R. (2012). Signals of change in tropical savanna woodland vertebrate fauna five years after feral herbivore removal. *Wildl. Res.*, in press.

Laffan S.W., Lubarsky E. & Rosauer D.F. (2010). Biodiverse, a tool for the spatial analysis of biological and related diversity. *Ecography*, 33, 643-647.

Laurie J., Mantle D. & Nicoll R.S. (2008). Customising the geological timescale. In: *AusGeo News 92*. Geoscience Australia, Australian Government Canberra, p. 3.

Leathwick J.R., Elith J., Francis M.P., Hastie T. & Taylor P. (2006). Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. *Marine Ecology Progress Series*, 321, 267-281.

Leibold M.A., Economo E.P. & Peres-Neto P. (2010). Metacommunity phylogenetics: separating the roles of environmental filters and historical biogeography. *Ecol. Lett.*, 13, 1290-1299.

Leibold M.A., Holyoak M., Mouquet N., Amarasekare P., Chase J.M., Hoopes M.F., Holt R.D., Shurin J.B., Law R., Tilman D., Loreau M. & Gonzalez A. (2004). The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett.*, 7, 601-613.

Lewis D. (2002). *Slower than the Eye Can See*. Tropical Savannas CRC, Darwin.

Lin H.S., McInnes K.J., Wilding L.P. & Hallmark C.T. (1999). Effects of soil morphology on hydraulic properties: II. Hydraulic pedotransfer functions. *Soil Sci. Soc. Am. J.*, 63, 955-961.

Lindenmayer D.B., McIntyre S. & Fischer J. (2003). Birds in eucalypt and pine forests: landscape alteration and its implications for research models of faunal habitat use. *Biol. Conserv.*, 110, 45-53.

Liu S.F., Raymond O.L., Stewart A.J., Sweet I.P., Duggan M.B., Charlick C., Phillips D. & Retter A.J. (2006). *Surface geology of Australia 1:1,000,000 scale, Northern Territory [Digital Dataset]*. The Commonwealth of Australia, Geoscience Australia (http://www.ga.gov.au), Canberra.

Ludwig J., Eager R.W., Williams R.J. & Lowe L.M. (1999). Declines in vegetation patches, plant diversity, and grasshopper diversity near cattle watering-points in the Victoria River district, northern Australia. *The Rangeland Journal*, 21, 135-149.

Mackey B.G., Berry S., Hugh S., Ferrier S., Harwood T. & Williams K. (2012). Ecosystem greenspots: identifying potential drought, fire and climate change micro-refuges. *Ecological Applications*, 22, 1852-1864.

Manion G. (2013). *.NET Generalised Dissimilarity Modeller (GDM) version 3.00 (developmental - 0.25, 04/07/2013)*. NSW Department of Climate Change and Water, Armidale.

Margules C.R. & Pressey R.L. (2000). Systematic conservation planning. *Nature*, 405, 243-253.

Marnham J. & Morris P.A. (2003). A seamless digital regolith map of Western Australia: a potential resource for mineral exploration and environmental management. In: *Western Australia Geological Survey,*

Annual Review 2002–03. Department of Mines and Petroleum, Government of Western Australia Perth, pp. 27–33.

Marra G. & Wood S.N. (2012). Coverage Properties of Confidence Intervals for Generalized Additive Model Components. *Scandinavian Journal of Statistics*, 39, 53-74.

McBratney A.B., Santos M.L.M. & Minasny B. (2003). On digital soil mapping. *Geoderma*, 117, 3-52.

McKenzie N.J. & Hook J. (1992). Interpretations of the Atlas of Australian Soils. In: *CSIRO Division of Soils Technical Report 94/199*. CSIRO Division of Soils Canberra.

McKenzie N.J., Jacquier D.W., L.J. A. & Cresswell H.P. (2000a). Estimation of Soil Properties Using the Atlas of Australian Soils. In: *CSIRO Land and Water Technical Report 11/00*. CSIRO Land and Water Canberra.

McKenzie N.J., Jacquier D.W., Maschmedt D.J., Griffin E.A. & Brough D.M. (2005). The Australian Soil Resource Information System Technical specifications. In. National Committee on Soil and Terrain Information / Australian Collaborative Land Evaluation Program (www.asris.csiro.au) Canberra, p. 89.

McKenzie N.L., Gibson N., Keighery G.J. & Rolfe J.K. (2004). Patterns in the biodiversity of terrestrial environments in the Western Australian wheatbelt. *Records of the Western Australian Museum, Supplement*, 67, 293–335.

Mckenzie N.L., Keighery G.J., Gibson N. & Rolfe J.K. (2000b). Patterns in the biodiversity of terrestrial environments in the southern Carnarvon Basin, Western Australia. *Rec West Aust Mus Suppl*, 61, 511–546.

McKenzie N.L., Leeuwen S.v. & Pinder A.M. (2009). Introduction to the Pilbara Biodiversity Survey, 2002–2007. *Records of the Western Australian Museum, Supplement*, 78, 3–89.

Milligan P. (2010a). Fifth Edition Total Magnetic Intensity (TMI) Anomaly Grid of Australia In. Geoscience Australia Canberra, p. 5.

Milligan P. (2010b). New magnetic datasets to identify energy, geothermal and mineral resources. New compilation of Magnetic Anomaly Map of Australia released. *AusGeo News*, Sept 2010, 1-2.

Minty B., Franklin R., Milligan P., Richardson L.M. & Wilford J. (2009). The Radiometric Map of Australia. *Exploration Geophysics*, 40, 325-333.

Mokany K., Harwood T.D., Williams K.J. & Ferrier S. (2012). Dynamic macroecology and the future for biodiversity. *Global Change Biology*, 18, 3149–3159.

Mott B., Alford R.A. & Schwarzkopf L. (2010). Tropical reptiles in pine forests: Assemblage responses to plantations and plantation management by burning. *For. Ecol. Manag.*, 259, 916-925.

Muir J., Schmidt M., TIndall D., Trevithick R., Scarth P. & Stewart J. (2011). Guidelines for Field measurement of fractional ground cover: a technical handbook supporting the Australian collaborative land use and management program. In: (ed. Management QDoEaR). Australian Bureau of Agricultural and Resource Economics and Sciences Canberra.

Nettleton L.L. (1971). Sources of Magnetic Anomalies and Interpretation. In: *Elementary Gravity and Magnetics for Geologists and Seismologists, Part II The Magnetic Method* (ed. Nettleton LL). Society of Exploration Geophysicists, p. 88.

NRS (2010). *Collaborative Australian Protected Areas Database - CAPAD 2008 (digital spatial data)*. National Reserve System - Parks Australia, Department of Sustainability, Environment, Water, Population and Communities: Australian Government, Canberra.

Percival P.J. (2010). *Index of Airborne Geophysical Surveys (Eleventh Edition)*. Geoscience Australia, Record 2010/13, Canberra.

Perry J., Fisher A. & Palmer C. (2011a). Status and habitat of the Carpentarian Grasswren (*Amytornis dorotheae*) in the Northern Territory. *Emu*, 111, 155-161.

Perry J., Williams K.J., Ferrier S. & Harwood T. (2013a). Interim biodiversity habitat condition layer, Pilbara Region. v1. Data Collection. In. CSIRO Data Access Portal https://www.data.csiro.au/dap/landingpage?pid=csiro:7652.

Perry J., Williams K.J., Ferrier S. & Harwood T. (2013b). Mean percent bare ground cover for the Pilbara region (2000-2011). v1. Data Collection. In. CSIRO Data Access Portal https://www.data.csiro.au/dap/landingpage?pid=csiro:7653.

Perry J.J., Kutt A. S., Garnett S. T., Crowley G. M., Vanderduys E. P. & Perkins G. C. (2011b). Changes in the avifauna of Cape York Peninsula over a period of 9 years: the relative effects of fire, vegetation type and climate. *Emu*, 111, 120-131.

Pierce D. (2013). Package 'ncdf'. In: *Interface to Unidata netCDF data files* CRAN.

Pinder A.M., Halse S.A., Shiel R.J. & McRae J.M. (2009). An arid zone awash with diversity: patterns in the distribution of aquatic invertebrates in the Pilbara region of Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 205-246.

Potts J.M. & Elith J. (2006). Comparing species abundance models. *Ecol. Modell.*, 199, 153-163.

Pringle H.J.R. & Landsberg J. (2004). Predicting the distribution of livestock grazing pressure in rangelands. *Austral Ecology*, 29, 31-39.

Prins H.H.T. & van Langevelde F. (2008). *Resource Ecology: Spatial and Temporal Dynamics of Foraging*. Springer.

R Development Core Team (2009). R Project for Statistical Computing. In. http://www.r-project.org/.

Randall R.P. (2007). *The introduced flora of Australia and its weed status*. Cooperative Research Centre for Australian Weed Management, Adelaide, South Australia.

Raymond O.L., Liu S.F. & Kilgour P. (2007a). *Surface geology of Australia 1:1,000,000 scale, Tasmania - 3rd edition [Digital Dataset]*. Geoscience Australia (http://www.ga.gov.au), Canberra.

Raymond O.L., Liu S.F., Kilgour P., Retter A.J. & Connolly D.P. (2007b). *Surface Geology of Australia 1:1,000,000 scale, Victoria - 3rd edition [Digital Dataset]*. Geoscience Australia (http://www.ga.gov.au), Canberra.

Raymond O.L., Liu S.F., Kilgour P., Retter A.J., Stewart A.J. & Stewart G. (2007c). *Surface geology of Australia 1:1,000,000 scale, New South Wales - 2nd edition [Digital Dataset]* The Commonwealth of Australia, Geoscience Australia (http://www.ga.gov.au), Canberra.

Rosauer D. (2009). *GDM Beta-Diversity Site-Pair Generator module for Biodiverse (Perl scripts)*. CSIRO Plant Industry, Canberra.

Rosauer D.F., Ferrier S., Williams K.J., Manion G., Keogh J.S. & Laffan S.W. (2013). Phylogenetic generalised dissimilarity modelling: a new approach to analysing and predicting spatial turnover in the phylogenetic composition of communities. *Ecography*, no-no.

Russell-Smith J., Whitehead P. & Cooke P. (2009). *Culture, Ecology and Economy of Fire Mangement in North Australian Savannas, Rekindling the Wurrk Tradition*.

Rybicki J. & Hanski I. (2013). Species–area relationships and extinctions caused by habitat loss and fragmentation. *Ecol. Lett.*, 16, 27-38.

Sinervo B., Mendez-de-la-Cruz F., Miles D.B., Heulin B., Bastiaans E., Cruz M.V.S., Lara-Resendiz R., Martinez-Mendez N., Calderon-Espinosa M.L., Meza-Lazaro R.N., Gadsden H., Avila L.J., Morando M., De la Riva I.J., Sepulveda P.V., Rocha C.F.D., Ibarguengoytia N., Puntriano C.A., Massot M., Lepetz V., Oksanen T.A., Chapple D.G., Bauer A.M., Branch W.R., Clobert J. & Sites J.W. (2010). Erosion of lizard diversity by climate change and altered thermal niches. *Science*, 328, 894-899.

Speight J.G. (2009). Landform. In: *Australian soil and land survey field handbook (3rd edn)* (ed. National Committee on Soil and Terrain). CSIRO Publishing Melbourne.

Stewart A.J., Sweet I.P., Needham R.S., Raymond O.L., Whitaker A.J., Liu S.F., Phillips D., Retter A.J., Connolly D.P. & Stewart G. (2008). *Surface geology of Australia 1:1,000,000 scale, Western Australia [Digital Dataset]*. The Commonwealth of Australia, Geoscience Australia (http://www.ga.gov.au), Canberra.

Turak E., Ferrier S., Barrett T., Mesley E., Drielsma M., Manion G., Doyle G., Stein J. & Gordon G. (2010). Planning for the persistence of river biodiversity: exploring alternative futures using process-based models. *Freshwater Biology*, no-no.

Urban M.C. (2004). Disturbance heterogeneity determines freshwater metacommunity structure. *Ecology*, 85, 2971-2978.

USGS (2013). Earth Resources Observation and Science Center [Web Page]. URL http://glovis.usgs.gov/

Van Vreeswyk A.M.E., Payne A.L. & Leighton K.A. (2004a). Pastoral resources and their management in the Pilbara region of Western Australia. In: (ed. Australia DoAW). Department of Agricutlure Perth.

Van Vreeswyk A.M.E., Payne A.L., Leighton K.A. & Hennig P. (2004b). An iventory and condition survey of the Pilbara region, Western Australia. In: (ed. Australia DoAW). Department of Agriculture South Perth, pp. 1-424.

VanDerWal J., Falconi L., Januchowski S., Shoo L. & Storlie C. (2012). SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exersises. In.

Viscarra-Rossel R.A., Bui E.N., de Caritat P. & McKenzie N.J. (2010a). Mapping iron oxides and the color of Australian soil using visible-near-infrared reflectance spectra. *J. Geophys. Res.*, 115, F04031.

Viscarra-Rossel R.A., Chappell A., Caritat P.d. & McKenzie N.J. (2010b). Mapping the information content of Australian visible-near infrared soil spectra In: *19th World Congress of Soil Science, Soil Solutions for a Changing World*. International Union of Soil Scientists 1–6 August 2010, Brisbane, Australia.

Viscarra-Rossel R.A. & Chen C. (2011). Digitally mapping the information content of visible-near infrared spectra of surficial Australian soils. *Remote Sens. Environ.*, 115, 1443-1455.

Viscarra Rossel R.A. (2011). Fine-resolution multiscale mapping of clay minerals in Australian soils measured with near infrared spectra. *J. Geophys. Res.*, 116, F04023.

Volschenk E.S., Burbidge A.H., Durrant B.J. & Harvey M.S. (2009). Spatial distribution patterns of scorpions (Scorpiones) in the arid Pilbara region of Western Australia. *Records of the Western Australian Museum, Supplement*, 78, 271-284.

Wallace J.F., Caccetta P.A. & Kiiveri H.T. (2004). Recent developments in analysis of spatial and temporal data for landscape qualities and monitoring. *Austral. Ecol.*, 29, 100-107.

Watson I.W., Novelly P.E. & Thomas P.W.E. (2007a). Monitoring changes in pastoral rangelands - the Western Australian Rangeland Monitoring System (WARMS). *Rangeland Journal*, 29, 191-205.

Watson I.W., Novelly P.E. & Thomas P.W.E. (2007b). Monitoring changes in pastoral rangelands â€" the Western Australian Rangeland Monitoring System (WARMS). *The Rangeland Journal*, 29, 191-205.

Western A. & McKenzie N. (2004). *Soil hydrological properties of Australia Version 1.0.1*. CRC for Catchment Hydrology, Melbourne.

Whitaker A.J., Champion D.C., Sweet I.P., Kilgour P. & Connolly D.P. (2007). *Surface geology of Australia 1:1,000,000 scale, Queensland - 2nd edition [Digital Dataset]*. Canberra: The Commonwealth of Australia, Geoscience Australia (http:/www.ga.gov.au).

Whitaker A.J., H.D. G., English P.M., Stewart A.J., Retter A.J., Connolly D.P., Stewart G.A. & Fisher C.L. (2008). *Surface geology of Australia 1:1,000,000 scale, South Australia - 1st edition [Digital Dataset]*. The Commonwealth of Australia, Geoscience Australia (http://www.ga.gov.au), Canberra.

Wilford J. (2012). A weathering intensity index for the Australian continent using airborne gamma-ray spectrometry and digital terrain analysis. *Geoderma*, 183-184, 124-142.

Williams K.J., Belbin L., Austin M.P., Stein J. & Ferrier S. (2012). Which environmental variables should I use in my biodiversity model? *International Journal of Geographic Information Sciences*, 26, 2009-2047.

Williams K.J., Ferrier S., Perkins G., Manion G., Harwood T. & Perry J. (2013a). GDM-scaled environmental predictors for compositional turnover in five biological groups across the Pilbara. v1. Data Collection. In. CSIRO Data Access Portal https://www.data.csiro.au/dap/landingpage?pid=csiro:7634.

Williams K.J., Ferrier S., Rosauer D., Yeates D., Manion G., Harwood T., Stein J., Faith D.P., Laity T. & Whalen A. (2010a). *Harnessing Continent-Wide Biodiversity Datasets for Prioritising National Conservation Investment*. A report prepared for the Department of Sustainability, Environment, Water, Population and Communities, Australian Government, Canberra, by CSIRO Ecosystem Sciences, Canberra.

Williams K.J., Ferrier S., Rosauer D., Yeates D., Manion G., Harwood T., Stein J., Faith D.P., Laity T. & Whalen A. (2010b). *Harnessing Continent-Wide Biodiversity Datasets for Prioritising National Conservation Investment: Appendices*. A report prepared for the Department of Sustainability, Environment, Water, Population and Communities, Australian Government, Canberra, by CSIRO Ecosystem Sciences, Canberra.

Williams K.J., Liu Y., Henderson B., Ferrier S., Perkins G. & Harwood T. (2013b). Predicted richness and standard errors for five biological groups across the Pilbara. v1. Data Collection. In. CSIRO Data Access Portal https://www.data.csiro.au/dap/landingpage?pid=csiro:7635.

Wilson J. & Gallant J. (2000a). Digital terrain analysis. In: *Terrain Analysis: Principles and Applications* (eds. JP W & JC G). John Wiley New York, pp. 1–27.

Wilson J.P. & Gallant J.C. (2000b). Secondary topographic attributes. In: *Terrain Analysis: Principles and Applications* (eds. Wilson JP & Gallant JC). John Wiley & Sons New York, pp. 51–85.

Woinarski J., Mackey B., Nix H. & Traill B. (2007). *The Nature of Northern Australia:Natural values, ecological processes and future prospects*. ANU E Press, Canberra.

Woinarski J., Risler J. & Kean L. (2004). Response of vegetation and vertebrate fauna to 23 years of fire exclusion in a tropical *Eucalyptus* open forest, Northern Territory, Australia. *Austral. Ecol.*, 29, 156-176.

Woinarski J.C.Z., Legge S., Fitzsimons J.A., Traill B.J., Burbidge A.A., Fisher A., Firth R.S.C., Gordon I.J., Griffiths A.D., Johnson C.N., McKenzie N.L., Palmer C., Radford I., Rankmore B., Ritchie E.G., Ward S. & Ziembicki M. (2011). The disappearing mammal fauna of northern Australia: context, cause, and response. *Conservation Letters*, DOI: 10.1111/j.1755-263X.2011.00164.x.

Wood S. (2013). R Package 'mgcv': Mixed GAM Computation Vehicle with GCV/AIC/REML smoothness estimation. In. Cran-R Project online, http://cran.r-project.org/web/packages/mgcv/index.html, p. 218.

Wood S.N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Xu T. & Hutchinson M. (2011). *ANUCLIM Version 6.1 User Guide*. The Australian National University, Fenner School of Environment and Society, Canberra.

Yee T.W. (2013). R Package 'VGAM': e Vector Generalized Linear and Additive Models. In. Cran-R Project online, http://cran.r-project.org/web/packages/VGAM/index.html, p. 783.

Zerger A., Williams K.J., Nicholls M., Belbin, Harwood T., Bordas V., Ferrier S. & Perkins G. (2013). *Biodiversity Profiling: Components of a continental biodiversity information capability*. Environmental Information Program Publication Series, No. 2, Bureau of Meteorology, Canberra, Australia.

Zuur A.F., Saveliev A.A. & Ieno E.N. (2012). *Zero Inflated Models and Generalized Linear Mixed Models with R*. Highland Statistics.