The Australian National Rabbit Database: 50 yr of population monitoring of an invasive species

Emilie Roy-dufresne, Miguel Lurgi, Stuart Brown, Konstans Wells, Brian Cooke, Greg Mutze, David Peacock, Phill Cassey, Dave Berman, Barry Brook, et al.

The Australian National Rabbit Database: 50 years of population monitoring of an invasive species

ROY-DUFRESNE, EMILIE [1,29], MIGUEL LURGI [1,2], STUART C. BROWN [1], KONSTANS WELLS [1,3], BRIAN COOKE [4], GREG MUTZE [5], DAVID PEACOCK [5,6], PHILL CASSEY [1], DAVE BERMAN [7], BARRY W. BROOK [1,8], SUSAN CAMPBELL [9], TARNYA COX [10], JOANNE DALY [11], IAIN DUNK [12], PETER ELSWORTH [13], DON FLETCHER [14], DAVID M. FORSYTH [10,15], GREG HOCKING [16,17], JOHN KOVALISKI [5], MICHAEL LEANE [18], BILL LOW [19], MALCOLM KENNEDY [20], JOHN MATTHEWS [21], STEVE MCPHEE [15], CAMILLE MELLIN [1,22], TRISH MOONEY [12], KATHERINE MOSEBY [23], JOHN READ [1], BARRY J. RICHARDSON [24], KATHRYN SCHNEIDER [25], ERIC SCHWARZ [16], RONALD SINCLAIR [5], TANJA STRIVE [26], FRANK TRIULCIO [27], PETER WEST [10], FREDERIK SALTRÉ [1,28], AND DAMIEN A. FORDHAM [1]

[1] The School of Biological Sciences, University of Adelaide, SA 5005, Australia
[2] Centre for Biodiversity Theory and Modelling. Theoretical and Experimental Ecology Station, French National Center of Scientific Research (CNRS) - Paul Sabatier University, 09200 Moulis, France
[3] Department of Biosciences, Swansea University, SA2 8PP, Wales, UK
[4] Institute for Applied Ecology, University of Canberra, ACT 2617, Canberra, Australia
[5] Biosecurity SA, Department of Primary Industries and Regions South Australia, SA 5064, Adelaide, Australia
[6] School of Animal and Veterinary Sciences, University of Adelaide, SA 5005, Adelaide, Australia
[7] University of Southern Queensland, QLD 4350, Toowoomba, Australia
[8] School of Natural Sciences, Private Bag 55, University of Tasmania, TAS 7001, Hobart, Australia
[9] Department of Primary Industries and Regional Development, Government of Western Australia, WA 6330, Albany, Australia
[10] Vertebrate Pest Research Unit, New South Wales Department of Primary Industries, NSW 2800, Orange, Australia
[11] Commonwealth Scientific and Industrial Research Organisation (CSIRO) Agriculture and Food, ACT 2601, Canberra, Australia
[12] Department of Environment, Water, and Natural Resources, Government of South Australia, SA 5000, Adelaide, Australia
[13] Biosecurity Queensland, Department of Agriculture and Fisheries, QLD 4350, Toowoomba, Australia
[14] Department of Environment and Planning Directorate, Australian Capital Territory Government, ACT 2602, Canberra, Australia
[15] Arthur Rylah Institute for Environmental Research, Department of Environment, Land, Water and Planning, VIC 3084, Melbourne, Australia
[16] Department of Primary Industries, Parks, Water and Environment, Government of Tasmania, TAS 7001, Hobart, Australia
[17] Agricultural Technical Services P/L, SA 5576, Yorketown, Australia
[18] Riverina local land service, NSW 2722, Gundagai, Australia
[19] Low Ecological Services, NT 0871, Alice Springs, Australia
[20] Department of Primary Industries and Regional Development, Government of Western Australia, WA 6151, South Perth, Australia
[21] Agricultural Services and Biosecurity Operations Division, Department of Economic Development, Jobs, Training and Resources, Goverment of Victoria, VIC 3300, Hamilton, Australia
[22] Australian Institute of Marine Science, QLD 4810, Townsville, Australia
[23] Centre for Ecosystem Science, University of New South Wales, NSW 2052, Sydney, Australia
[24] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Austrialian National Insect Collection (ANIC), and National Research Collections Australia (NRCA), ACT 2601, Canberra, Australia
[25] Parks Victoria, VIC 3000, Melbourne, Australia
[26] Commonwealth Scientific and Industrial Research Organisation (CSIRO) Health and Biosecurity, ACT 2601, Canberra, Australia
[27] Department of Planning, Transport and Infrastructure, Government of South Australia, SA 5000, Adelaide, Australia
[28] Global Ecology, College of Science and Engineering, Flinders University GPO Box 2100, SA 5001, Adelaide, Australia

[29] E-mail: roydufresne.emilie@gmail.com

INTRODUCTION

Many species have been historically translocated between continents as domestic animals (Anderson 2009), ornaments (Reichard & White 2011), or for leisure activities (Spencer & Hampton 2005). During the 18^{th} and 19^{th} centuries, the European rabbit (*Oryctolagus cuniculus*) was introduced into Australia primarily as a food resource and for hunting activities (Munday 2017). The species established and spread rapidly across the continent due to its high reproductive capacity and ability to exploit a diversity of grassland ecosystems. Today, the European rabbit occupies >70% of the Australian continent (Stodart & Parer 1988), suppressing the regeneration of native vegetation (Forsyth *et al.* 2015), competing with native and domestic mammals for food and habitat (Moseby *et al.* 2009, Bird *et al.* 2012), sustaining inflated populations of invasive predators (Lurgi *et al.* 2018, but see Scroggie *et al.* 2018), and causing approximately A$200 million per year of associated economic losses (Gong 2009, Cooke 2013). Consequently, there has and continues to be a strong imperative to resolve the problem through active management.

Efforts to find a solution to the problem caused by the European rabbit in Australia began in the 1880s with the establishment of the Intercolonial Rabbit Commission and the adoption of various rabbit monitoring measures (Fenner & Fantini 1999). With the introduction of the viral disease myxomatosis in 1950, the Wildlife Section of CSIR (now CSIRO) led researchers on rabbit population control backed by state governmental research groups which adapted its directives to meet local requirement (Cooke, *pers. comm.*). The government at state and federal levels, industry, community groups, and landholders invested large amounts of resources and time into managing and controlling rabbit populations to limit their various effects on Australian ecosystems and agricultural industries. In time however, the CSIRO withdrew its lead role, leaving the state research groups to work increasingly as a quasi-national team (Cooke 2018) but with a tendency for actions to be based on a local understanding of drivers of rabbit abundance (Bowen & Read 1998, Scanlan *et al.* 2006, Fordham *et al.* 2012). A lack of adequate resources to manage and share data between researchers and government agencies has prevented systematic analyses of the ecological processes driving spatiotemporal variation of rabbit populations at appropriate spatial scales for comprehensive management (National Land & Water Resources Audit and Invasive Animals Cooperative Research Centre 2008). It is therefore critical to develop a single and harmonized database framework to collate rabbit occurrence and abundance data.

Here we provide the *Australian National Rabbit Database* (ANRD) which combines over 50 years of rabbit occurrence and abundance survey data from across Australia. The database collates information from > 120 studies across all states and territory administrative divisions of the Commonwealth of Australia into a rigorous, consistent, and unified framework, eliminating potential disparities resulting from the multiple monitoring and reporting methods used for data collection. This is one of the largest compilations of data on an invasive species, providing an important resource for determining drivers of patterns of variation in rabbit occupancy, abundance and population growth at spatial scales ranging from local to continental (Figure 1). The database complements previously published information on rabbit biology by providing the data required to quantify the role of climatic and environmental processes on the population ecology of the rabbit in Australia. Such information can be used to address important questions regarding the ecological determinants of range limits, the mechanisms that govern heterogeneity in spatial abundance patterns, the role of density dependence and environmental stochasticity in regulating life history traits, and the role of the species in its host ecosystem by considering the importance of inter-species relationships.

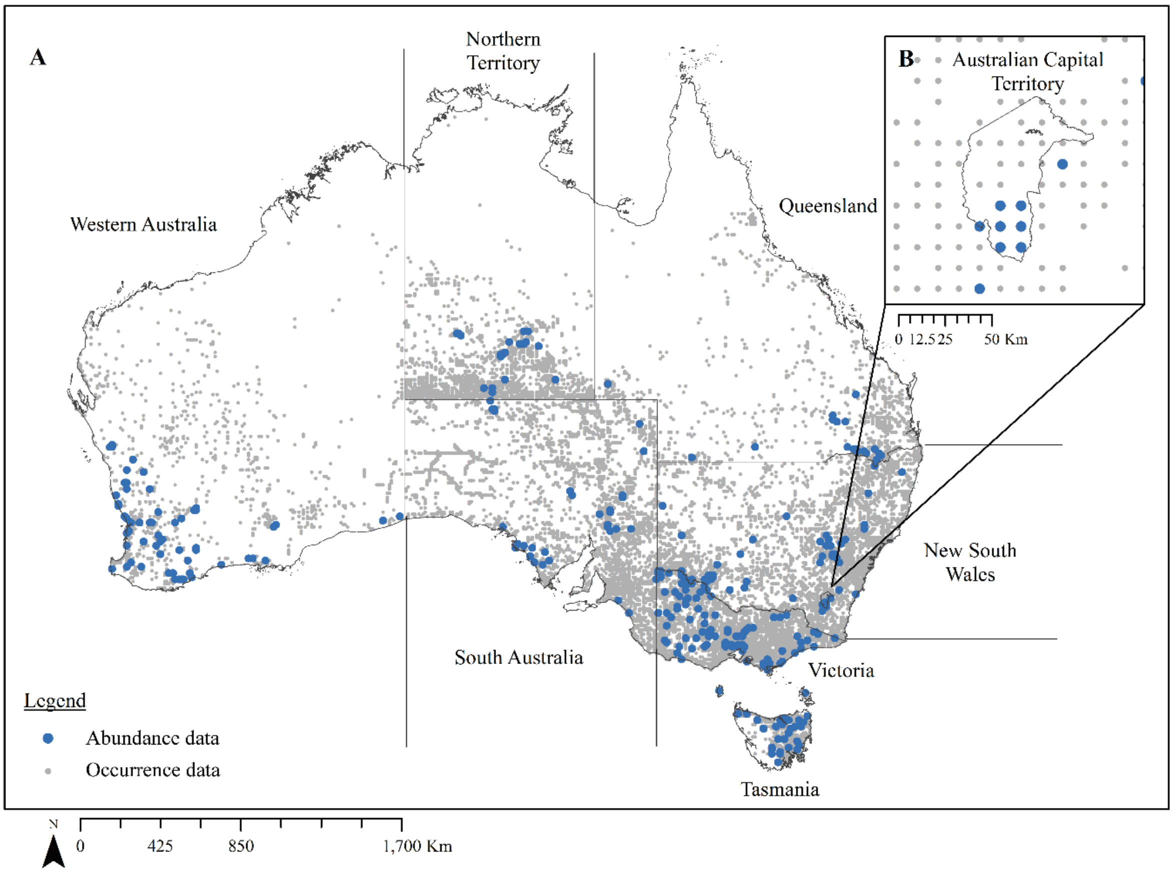

Figure 1: Spatial extent and distribution of the data provided within the rabbit database at the scale

of (A) Australia, showing state boundaries in black and (B) a close-up view of the Australian Capital Territory.

METADATA

CLASS I. DATA SET DESCRIPTORS

A. **Data set identity:**

   Title: The Australian National Rabbit Database: 50 years of population monitoring of an invasive species.

B. **Data set identification code:**

   AustralianNationalRabbitDatabase_OccurrenceData.csv

   AustralianNationalRabbitDatabase_AbundanceData.csv

C. **Data set description:**

   1. **Originators:**

      Emilie Roy-Dufresne

      The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

      Dr Damien Fordham

      The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

      Abstract:

      With ongoing introductions into Australia since the 1700s, the European rabbit (*Oryctolagus cuniculus*) has become one of the most widely distributed and abundant vertebrate pests, adversely impacting Australia's biodiversity and agro-economy. To better understand the population and range dynamics of the species and its impacts, occurrence and abundance have been collected by researchers and citizens from sites covering a broad spectrum of climatic and environmental conditions in Australia. The lack of a common and accessible repository for these data has, however, limited their

use in determining important spatiotemporal drivers of the structure and dynamics of the geographical range of rabbits in Australia. To meet this need, we created the Australian National Rabbit Database which combines more than 50 years of historical and contemporary survey data collected from throughout the range of the species in Australia. The survey data, obtained from a suite of complementary monitoring methods, were combined with high-resolution weather, climate and environmental information, and an assessment of data quality. The database provides records of rabbit occurrence (689,265 records) and abundance (51,241 records, > 120 distinct sites) suitable for identifying the spatiotemporal drivers of the rabbit's distribution and for determining spatial patterns of variation in its key life history traits, including maximum rates of population growth. Since all data are georeferenced and date stamped, they can be coupled with information from other databases and spatial layers to explore the potential effects of rabbit occurrence and abundance on Australia's native wildlife and agricultural production. The Australian National Rabbit Database is an important tool for understanding and managing the European rabbit in its invasive range and its effects on native biodiversity and agricultural production. It also provides a valuable resource for addressing questions related to the biology, success, and impacts of invasive species more generally. No copyright or proprietary restrictions are associated with the use of this data set other than citation of this Data Paper.

D. Key words:

European rabbit, *Oryctolagus cuniculus*, long-term monitoring data, invasive species management, occupancy, population abundance, demography, weather, long-term climatic data.

E. Descriptions:

Occurrence records: European rabbit occurrence records (689,265 entries over 9,839 10 km^2 grids) were collected from a variety of sources (156 studies; Table 1; Supplementary Information SI1, also see Supplementary Information SI3 for a list of the studies), including: (*i*) sporadic observations from expert (6.82%) and citizen scientists (i.e. Feral Scan Data - https://www.feralscan.org.au/; 0.74 %), (*ii*) abundance monitoring programs such as spotlight surveys (0.12 %), (*iii*) signs of presence (0.07%), (*iv*) warrens locations (89.24 %), (*v*) rabbit sampling such as live-trapping sites and shootings (2.63 %), and (*vi*) vector and disease release and monitoring activities (0.38%). The data cover the years between 1760 and 2015, although most occurrences (>99%) are reported between 1978 and 2015 (Figure 2). The occurrence data are spatially representative of the established distribution of the rabbit in Australia (Figure 3), comprising a variety of land systems, vegetation structures, and ecosystems, and including occurrences on both sides of the wild dog barrier, which stretches through the states of South Australia, New South Wales, and Queensland. For each record, information is provided (see Table 5 in section Class IV.B) at a 10 km^2 spatial resolution on its (*i*) geolocation (i.e. ID, coordinates, aggregated total number of records in the grid, geographical bias index, and the State where it is situated), (*ii*) source (i.e. data type, source ID and name, and site section name), (*iii*) the date of collection (i.e. year, month, and day), and (*iv*) a comprehensive assessment of its quality value including: accuracy and reliability, completeness, temporal coverage, and consistency.

Table 1: Summary of the information provided for the occurrence data.

State/territory	No. studies	Time period	No. years	Total no. grid with occurrence data (citizen)[1]	Total no. grid with occurrence data (expert)[2]
VIC	28	1760-2014	104	463	1,664
NSW	8	1825-2014	59	108	389
ACT	12	1866-2013	36	22	1,218
QLD	34	1837-2015	81	333	2,124
SA	18	1837-2014	42	491	336
NT	24	1850-2014	37	259	784
WA	22	1800-2014	78	708	2,201
TAS	10	1964-2014	30	27	46

[1] Citizen occurrence data were obtained from the Feral Scan Database (https://www.feralscan.org.au/).

[2] Expert occurrence data were obtained from expert monitoring observation and signs of presence, data from the Atlas of Living Australia (https://www.ala.org.au/), recorded warren locations,

shooting locations, spotlight transects locations, sites where Spanish fleas (*Xenopsylla cunicularis*) were released, and the survey sites for the RHDV1 spread.

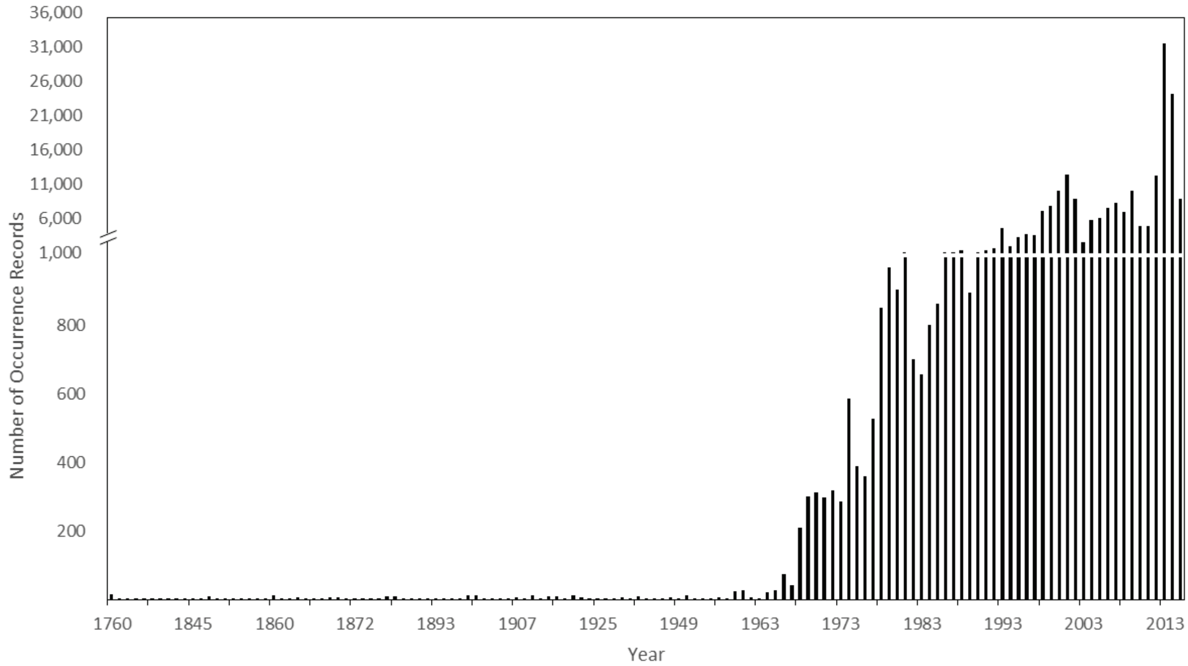

Figure 2: Number of occurrence entries per year. Only 335 entries (~0.05 % of the entries) were obtained between 1760 and 1959).

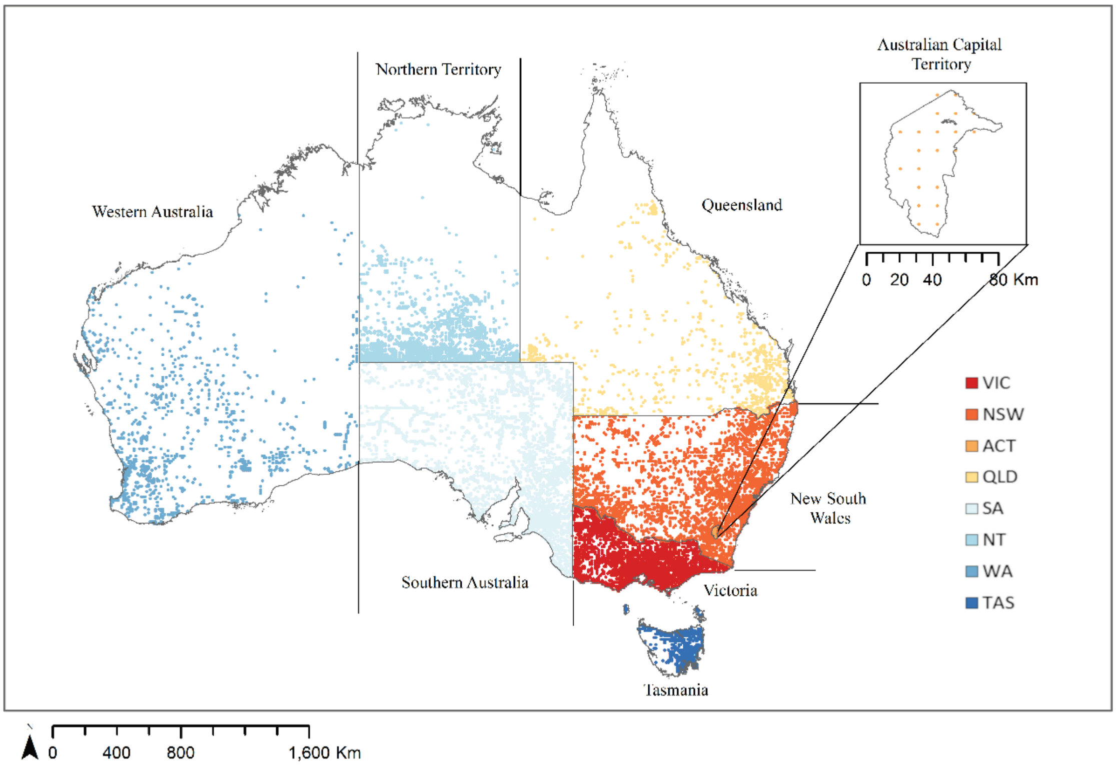

Figure 3: Spatial distribution of the occurrence data for each state and territory.

Abundance records: Rabbit population abundance measures is commonly based on spotlight surveys across Australia which provide an index of the abundance of adult individuals (i.e. >800 g., Cooke, 1970). Rabbits are counted along transects that are approximatively 10 km long and 50 or 100 m wide. Rabbit counts are converted into the number of individuals observed per km and repeated over 2-3 consecutive nights avoiding poor weather conditions. The spotlight data are reported as provided by the different studies and are not corrected for the species' probability of detection. They comprise a total of 51,241 records from 39 studies and 1,858 transects (i.e. 305 grids), covering the period from 1965 to 2015 (Table 2; Supplementary Information SI2, also see Supplementary Information SI4 for a list of the studies). Most of these transects were visited between 3 and 24 times over ten years (Figure 4). Spotlight data are provided at 10 km^2 grid cell resolution (Figure 5), and we included the following features: (*i*) geographic coordinates (*ii*) details of sampling design such as total number, lengths and width of transects in the grid, and the state where it is located), (*iii*) the source of the information (i.e. source ID and name, the ID of the study region, and site section name), (*iv*) the visit number, (*v*) the total number of replicates per visit, (*vi*) the date of collection (i.e. year, month, and day), (*vii*) the ID of the transect, (*viii*) information on the recorded abundance (i.e. raw number of rabbit per km), (*ix*) any additional notes (see hereafter), and (*x*) a compressive assessment of the record quality value, including: accuracy, completeness, temporal coverage, and consistency (see Table 7 in section Class IV.B). The 'additional notes' provides extra information provided by the data collector regarding technical aspects and relevant observations made at the time of data collection.

Table 2: Summary of the information provided for the abundance data.

State/territory	No. studies	Time period	No. years studied	Total no. transects
VIC	6	1970-2014	36	728
NSW	7	1970-2014	23	94
ACT	3	1995-2014	20	39
QLD	6	1971-2012	25	35
SA	17	1965-2015	51	87
NT	6	1980-2012	21	228
WA	10	1992-2014	17	460
TAS	2	1975-2012	38	188

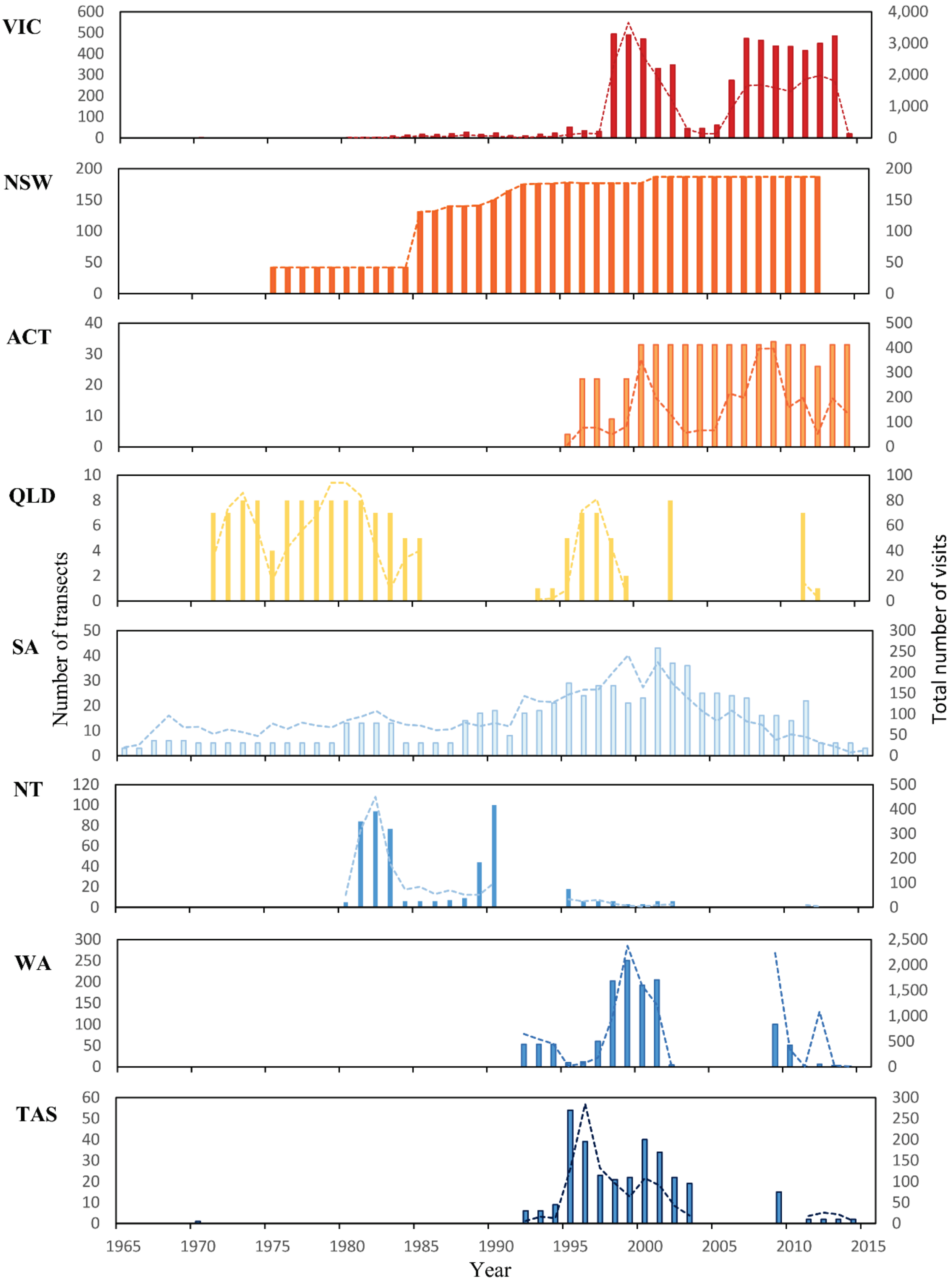

Figure 4: Number of transects (bars) and total number of visits summed across all transects (i.e. number of times the transects were visited; dashed lines) per state/territory and per year.

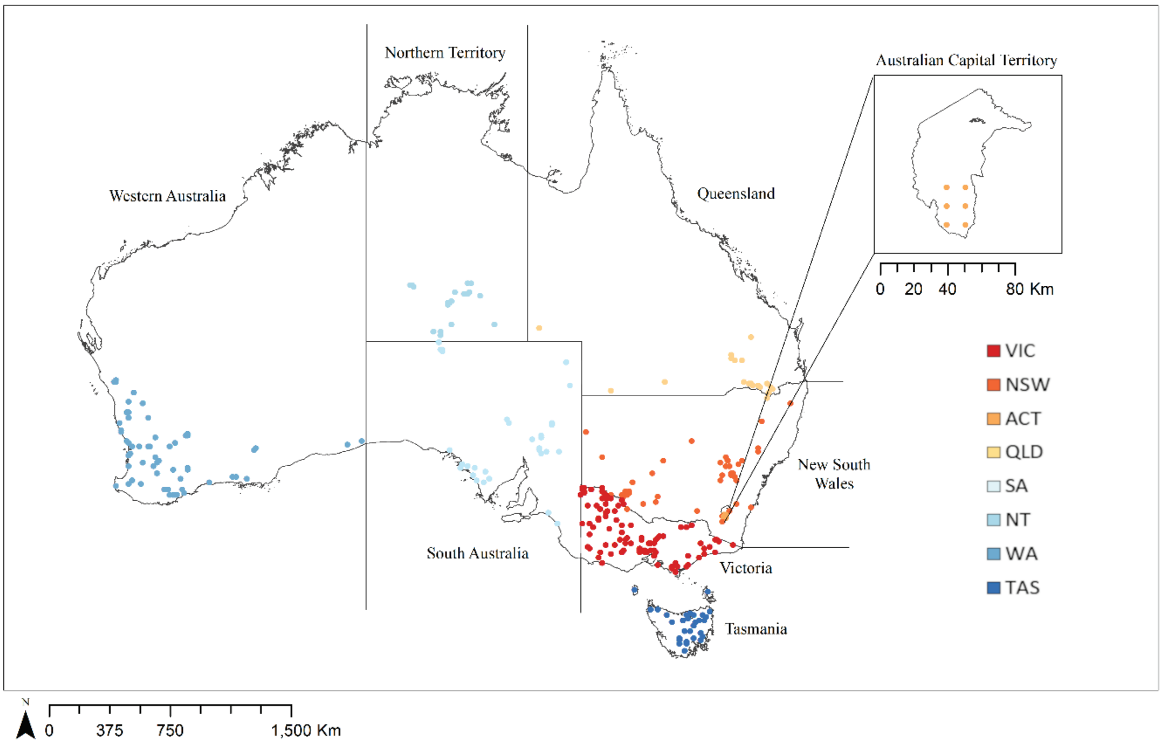

Figure 5: Spatial distribution of the spotlight count data of rabbit populations for each state/territory in Australia.

F. Research opportunity:

The database provides a prime opportunity to better understand the ecological processes that underpin the range and abundance of the European rabbit in Australia. Although the rabbit has been well-studied over the last 50 years in Australia, the database will help to develop process-based models to inform current pest management strategies, by providing vital information on variation in vital life history traits (e.g. Wells *et al.* 2016a). The database includes ecological information on rabbit distribution in time and space, including long-term time series abundance data, which inform fine-to-broad-scale predictions about the population dynamics of the rabbit (e.g. Scroggie *et al.* 2018) and its impacts. For example, the database is suitable for exploring how climate can influence the reproductive biology of invasive species (e.g. Wells *et al.* 2016b). It is also likely to be informative for ecological theories related to range size-abundance relationships, host-pathogen dynamics (e.g. Wells *et al.* 2015), meta-population dynamics and connectivity and the large-scale intra- and interspecific geographical patterns in population density (e.g. Lurgi *et al.* 2016), or the potential effects of rabbit removal on the whole ecological community (e.g. Pedler *et al.* 2016, Lurgi *et al.* 2018, Scroggie *et al.* 2018).

The database provides sampled and harmonized information across Australia over many decades. In doing so, it facilitates data sharing between State/territory agencies and individual researchers, providing new opportunities to conduct national scale research. By providing information on data quality (including measurement error), users can focus their analysis on the best quality data, or use hierarchical approaches to directly account for sources of bias and uncertainty in their analysis. The database and supporting metadata provides scope to develop consistent attributes to manage information on the species.

CLASS II. RESEARCH ORIGIN DESCRIPTORS

A. Overall project description:

1. **Identity:**

   An integrated tool for informing pest management: modelling range shifts for an invasive vertebrate in response to climate change.

2. **Originators:**

   Dr Damien Fordham

   The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

   Dr Phill Cassey

   The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

   Dr Barry Brook

   The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.
   (Faculty of Science, Engineering & Technology, Private Bag 55, University of Tasmania, TAS 7001, Australia.)

   Greg Mutze

   Biosecurity SA, Department of Primary Industries and Regions South Australia, SA 5064, Australia.

3. **Period of study:**

   The research project started in October 2013 and finished in December 2018. The data collection started in October 2013 and finished in December 2017.

4. **Objectives:**

   The project aims to develop an integrated pest management model that directly improves the effectiveness of control options for the European rabbit across their Australian range, including projections of future change. More specifically, the analysis and modelling approach will explore the relative efficacy of different rabbit pest management options and determine model sensitivities to assumptions model and a climate model with a spatially explicit demographic model.

5. **Abstract:**

   Invasive species and climate contribute directly to the loss of biodiversity and economic productivity. This research project focuses on providing user-oriented tools that enable a strategic approach to European rabbit management and vertebrate pest control in Australia in response to anticipated climate and land-use change.

6. **Source of funding:**

   The project was supported by the Australian Research Council (ARC; LP110200805 and FT140101192).

B. Specific subproject description:

1. **Site description:**

   a. **Geography:**

      The study area covers continental Australia, Tasmania, and the islands in Bass Strait (area bound by 112.92°E to 159.11°E and -43.74°S to -9.14°S).

   b. **Habitat:**

      Australia (excluding the sub-Antarctic regions) has seven ecoregions (Figure 6) (https://www.worldwildlife.org/biomes). The centre of Australia is dominated by desert and xeric shrubland habitats, and it is surrounded by Mediterranean forests and woodlands (along with the southern and western borders) and tropical and subtropical habitats (along with the northern border). Eastern Australia presents temperate types of habitats, including grasslands, savannas, broadleaf, and mixed forests with some regions in the south-eastern parts of Australia containing montane grasslands and shrubland habitats.

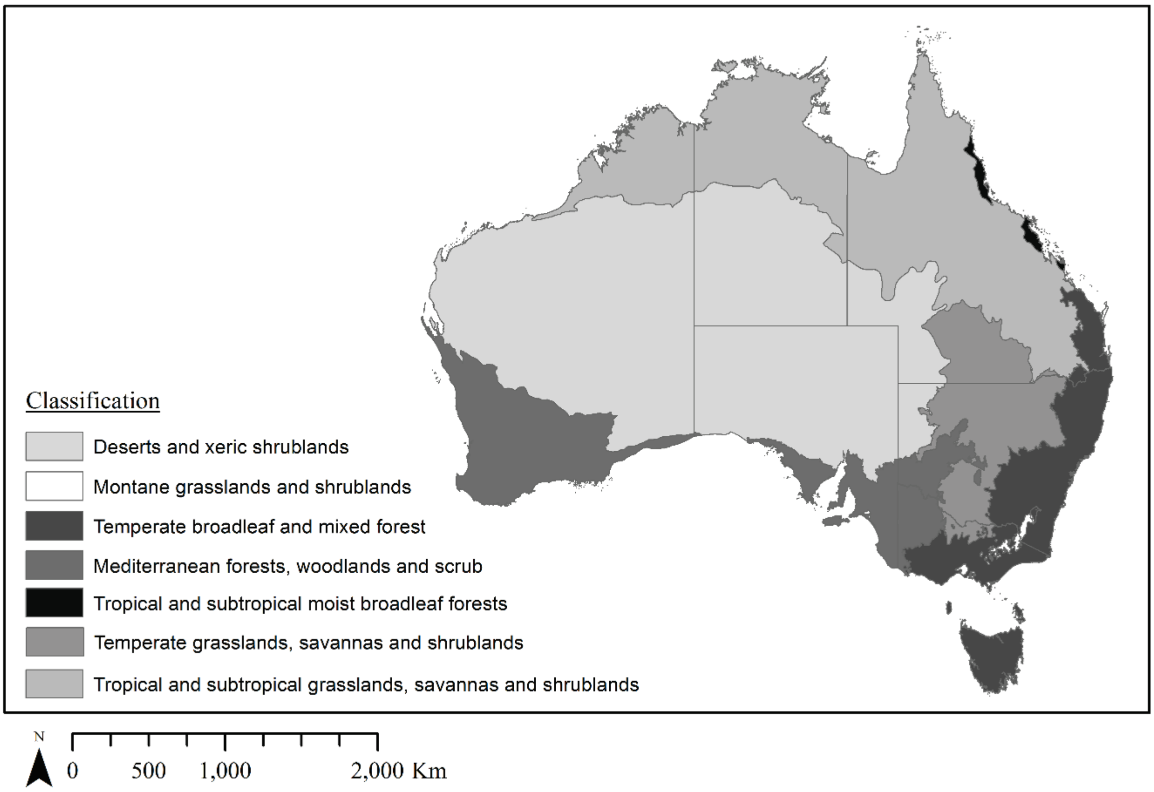

Figure 6: Australian ecoregions classification according to the terrestrial ecoregions classification established by the World Wildlife Fund (https://www.worldwildlife.org/biomes).

c. Geology and landform:

Australia (9° and 44°S latitude and 112° and 154°E longitude) is the flattest continent in the world with a landmass of 7,617,930 km^2 surrounded by coastlines (i.e. 34,218 km; Blewett 2012). It includes a large variety of biomes, such as tropical forests (north-east regions), mountain ranges (south-east, south-west, and eastern regions), and desert (centre regions, McKenzie & Ryan 1999). The soil is old and among the least fertile in the world (Bowler 1976), with the desert and semi-arid land covering 70% of the continent.

d. Site history:

The rabbit was first introduced to Australia in the 18th century by acclimatisation societies, which wanted to enrich the biodiversity of the country and enhance its hunting attractions (Williams *et al.* 1995). In 1827, the European rabbit became an issue in Tasmania, while further releases were occuring in mainland Australia. Deliberate introductions continued until 1880, after which rabbit density increased greatly (Thompson & King 1994). In response to these high rabbit densities, thousands of kilometres of rabbit-proof fences were erected, and control activities such as poisoning were conducted to halt further spread. The most successful control strategy was to rip the warrens with tractors, but this was time-consuming (Williams *et al.* 1995). The advent of myxomatosis in 1950, followed by the introduction of the rabbit haemorrhagic disease virus (RHDV) in 1995, decreased the rabbit population density across the continent, making traditional rabbit controls methods such as poisoning and warren ripping more effective (Henzell *et al*. 2002, 2008, Ramsey *et al.* 2014).

e. Climate:

Sea surface temperatures, drive most of the Australian climate creating periods of drought, seasonal tropical rainfalls (including cyclones), and important year-to-year rainfall variation (Ashok *et al.* 2003, England *et al.* 2006). Both tropical and equatorial climates characterise the northern part of the country, and humid subtropical conditions predominate in the eastern part (Figure 7). The southern and south-eastern parts of Australia undergo a temperate climate while semi-arid to arid conditions dominate the interior ranges (Peel *et al.* 2007).

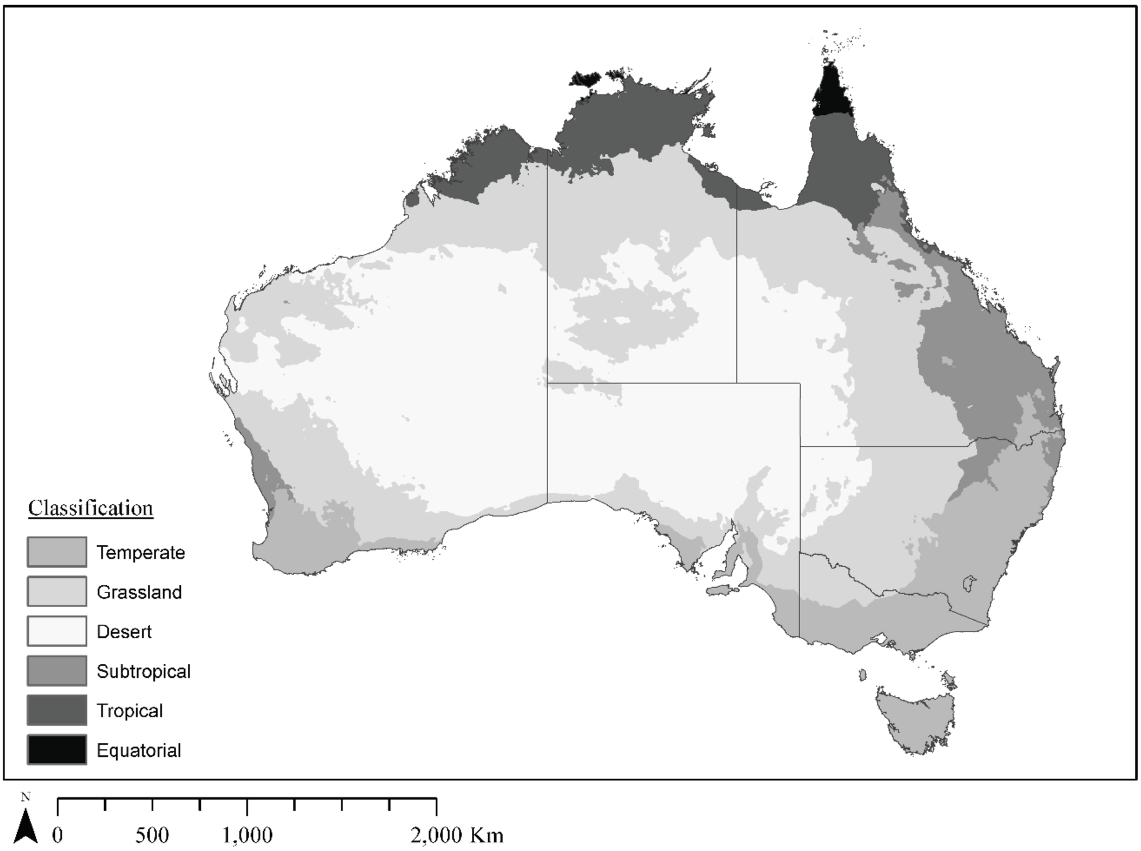

Figure 7: Distribution of climate zones across Australia according to the Köppen climate classification based on mean monthly and annual rainfall, maximum temperature and minimum temperature grids (http://www.bom.gov.au/climate/averages/climatology/gridded-data-info/metadata/md_koppen_classification.shtml).

2. **Experimental or sampling design (see Figure 8):**

   a. **Criteria for selection:**

   All data sources needed to meet three criteria to be included in the *Australian National Rabbit Database*. They must have (*i*) quantified rabbit occurrence or abundance data, (*ii*) provided the geolocality of the site along with the date of the sampling, information on the origin of the data, and the sampled transect (for abundance data), and (*iii*) specified the amount of sampling replication (if any).

   b. **Data collection:**

   We gathered the datasets from governmental agencies (state and territory) and relevant non-government sources, including researchers, environmental consultants and conservation organisations. These data included land management surveys and field sampling programs, and data obtained from the scientific literature by contacting corresponding authors. More data were obtained from discussions with monitoring contractors, landholders, and the broader community. We treated all types of data equally to maximise the size of the database while providing a quality assessment to document the reliability of the data collected.

3. Research methods (see Figure 8):

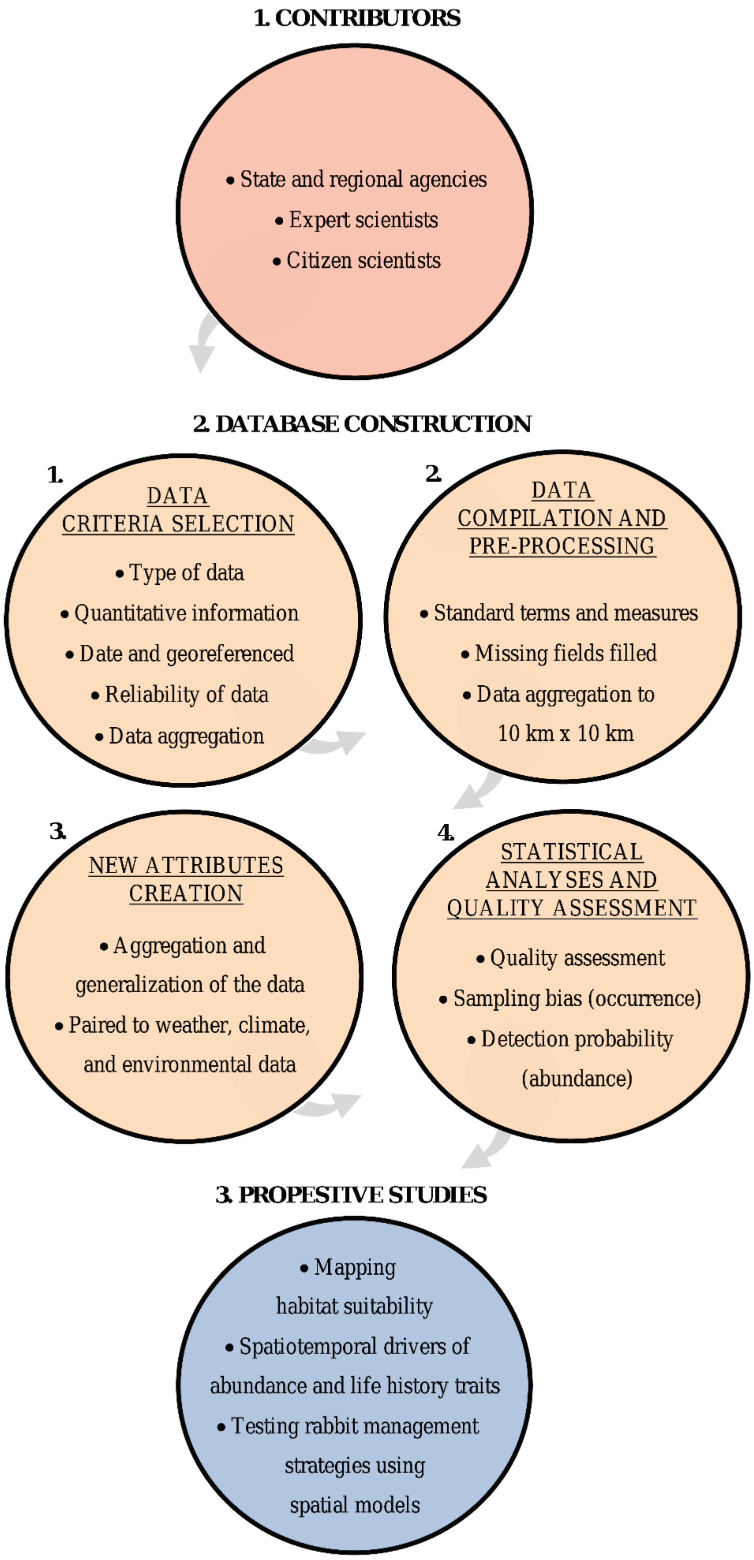

Figure 8: Flowchart diagram of the database. The colour code indicates the different phases of the project (i.e., contribution – in red, dataset construction – in yellow, and prospective studies- in blue) and the arrows inform the flow of the processed information.

a. **Data compilation:**

The data are compiled following a table format. There are two tables: one for occurrence data and one for abundance data. Each row of the tables contains a single and unique record. Geo-reference locations have been mapped to a 10 km^2 grid cell resolution to meet privacy concerns. The grid cell size was selected to represent the maximum natal dispersal distance for the rabbit, i.e., a radius of five kilometres (Twigg *et al.* 1998).

b. **Data pre-processing:**

Table 3 describes the steps followed to clean and transform the data to ensure a consistent format throughout the dataset, preventing issues of pseudo-replication and reducing potential sources of error. Data errors were addressed where possible (if any required information was accessible) or replaced by a 'NA' value in the database.

Table 3: Data pre-processing approach.

Step	Description
1. Format the data to match each field of the database.	Ensures that the required fields are present (i.e. geographical coordinates, data source name, site section name, data type for occurrence data, number of rabbits per kilometre for the abundance data). Format the fields so that they correspond to the ones selected for the database.
2. Convert the geographical coordinates to the format adopted in the database.	Identify the UTM grid corresponding to the location of the data and provide its centre point as geographical coordinates (WGS84 format). If the initial centre point falls into the water, use the location corresponding to a one km inland distance from the coastal segment the closest to the initial centre point.
3. Convert the recorded abundance values to the format adopted in the database.	Convert all abundance data into a standardised format corresponding to the number of rabbits km^{-1}.

4. Identify duplicate records to prevent pseudo-replication.	Remove duplicate records across the datasets.
5. Evaluate the reliability of the information for each record.	This includes traceability of the information (e.g. name of the site, evaluating that the geographic coordinates are documented, and the values are possible, verifying that a given date is possible, and confirming if the recorded abundance value is realistic) by confirming dubious information with the authors and other experts and referring to related publications, administrative boundaries, and online maps and imageries (see Hijmans *et al.* 1999, Yesson *et al.* 2007).

c. Creation of new attributes:

We generated new attributes for each feature to improve the querying process from the dataset (e.g. by aggregation and generalisation), and provide additional valuable information for prospective analyses (e.g. climatic and environmental data). The list and a short definition of the new attributes are provided below (Table 4). We also shared the scripts in the Supplementary Information SI5 and SI6.

Table 4: New attributes.

Attribute	Description
1. Australian state.	Name of the state/territory where the data were collected based on the location of the occurrence data or of the middle segment of the spotlight transect. We used state abbreviations: WA for Western Australia, NT for Northern Territory, SA for South Australia, QLD for Queensland, NSW for New South Wales, ACT for the Australian Capital Territory, VIC for Victoria, and TAS for Tasmania.
2. Total number of occurrences per UTM grid.	Only for the occurrence data, we calculated the total number of recorded occurrences falling into each 10 km^2 grid cell.
3. Total number of transects per UTM grid.	Only for the abundance data, we calculated the total number of transects falling into each grid cell of 10 km^2 spatial resolution.

4. Visit number.	Only for abundance data, we provided the visit number. It included the date when the data were collected and the total number of time the site was visited.
5. Total number of replicates.	Only for the abundance data, we provide the replicate number. We defined a replicate as any data sampled more than one time in less than seven days.

The occurrence and abundance data were also spatially and temporally matched to daily weather conditions, monthly mean and annual mean climate values, and 30-year mean climate and environmental data (see Table 5 and see description of the environmental variables, provenance, related processing tasks, in Section V.D). We provide 18 variables for the occurrence data (i.e. using 30-year mean climate and environmental data), and 38 variables for the abundance data (i.e. including, weather variables, mid-to-long term climate variables and environmental variables). These variables were selected based on expert knowledge of the climate and environmental factors that influence the occurrence and population dynamics of rabbits. Weather and climatic conditions were accessed through eMAST as part of the Terrestrial Ecosystem Research Network project (TERN project; Hutchinson *et al.* 2009, Hutchinson & Xu 2013, Hutchinson *et al.* 2014) at a 1 km^2 spatial resolution and upscaled to a 10 km^2 resolution by taking the mean for continuous variables and the mode for categorical variables.

Table 5: Summary of the weather, climate, and environmental variables matched to rabbit occurrence and abundance data. The climatic data were obtained from eMAST as part of the Terrestrial Ecosystem Research Network project (TERN project; Hutchinson *et al.* 2009, Hutchinson & Xu 2013, Hutchinson *et al.* 2014). Sources of information and data transformation associated with environmental variables are shown in table 8.

Variables	Occurrence data	Spotlight data

Daily values: Prec [1] (+1 month lag) TMax [2], TMin [3]		✓
Mean monthly values: Prec [1] (+ 1 and 2 yr lags, + 2 previous seasons*) TAvg [4] (+ previous summer and winter*)		✓
Mean annual values: Prec [1] (+ 1 and 2 yr lags) TMax Warmest Month [5] (+ 1 yr lag) TMin Coldest Month [6] (+ 1 yr lag) PSea [7] (+ 1 and 2 yr lags)		✓
30-year mean values: Prec [1], PSea [7], PAvg [8] (for every seasons*) TAvg [4], TMax [2], TMin [3], TSea [9], TAvg Wettest Month [10], Tavg Warmest Month [11], VegeType [12], DistPermWater [13], DistAgriLand [14], PercSoilClay [15], MinDayLength [16], VarDayLength [17]	✓	✓

[1] Total rainfall
[2] Maximum temperature
[3] Minimum temperature
[4] Average temperature
[5] Maximum temperature for the warmest month
[6] Minimum temperature for the coldest month
[7] Rainfall seasonality
[8] Average rainfall
[9] Temperature seasonality
[10] Average temperature of the wettest quarter
[11] Average temperature of the warmest quarter
[12] Major vegetation types
[13] Euclidean distance to permanent water features
[14] Euclidean distance to agricultural land margins
[15] Percentage of estimated clay
[16] Minimum day length
[17] Yearly variance in day length
* Following the Australian Calendar

d. Statistical analyses - geographical bias in sampling effort:

To estimate the geographical bias in the sampling effort for rabbit occurrence, we calculated the standardised density of occurrence data points (Phillips & Dudík 2008) using the function density.ppp from the Spatstat package in R (Baddeley *et al.* 2015; see Supplementary Information SI7). This function computes a Gaussian smoothing kernel value for the centre of each grid cell of 10 km^2 resolution and counts the number of points per unit area (i.e. 1 km^2). The points further away from the centre of the grids weight less on the final density of the grid cell than the point closer to the centre. We calculated two density values: one considering all the occurrence points and the second considering only the centre point of each 10 km^2 grid which includes an occurrence point. The former indicates the data density for the complete dataset including potential replicates within each 10 km^2 grid cell, whereas the latter eliminates any influence of replicates within each grid cell which could be biased by the nature of the occurrence data collected (e.g. warren ripping programs). The two metrics were standardised from 0 to 1 and can be interpreted as occurrence intensity estimates (i.e. the number of observations expected to be randomly collected in each grid given the current spatial pattern observed in the occurrence data, value going from 0 to 1; Figure 9). They give indications of regions which can present a high probability of being affected by geographical bias associated with uneven sampling effort resulting from data collected in a subjective manner, e.g. citizen science data collected during recreational activities in areas with high natural interest (Fourcade *et al.* 2014), or the process of collating together disparate datasets collected following different research objectives (Hortal *et al.* 2008, Robertson & Barker 2006). Presence of geographical bias is important to consider in the modelling parameters, as it can artificially inflate the probability of the presence of a species in a local area, making the data not representative of any ecological process but rather an artefact of the sampling process (Pocock *et al.* 2017). Geographical bias leads to erroneous interpretations from the model predictions (Dennis *et al.* 2006, Osborne & Leitão 2009).

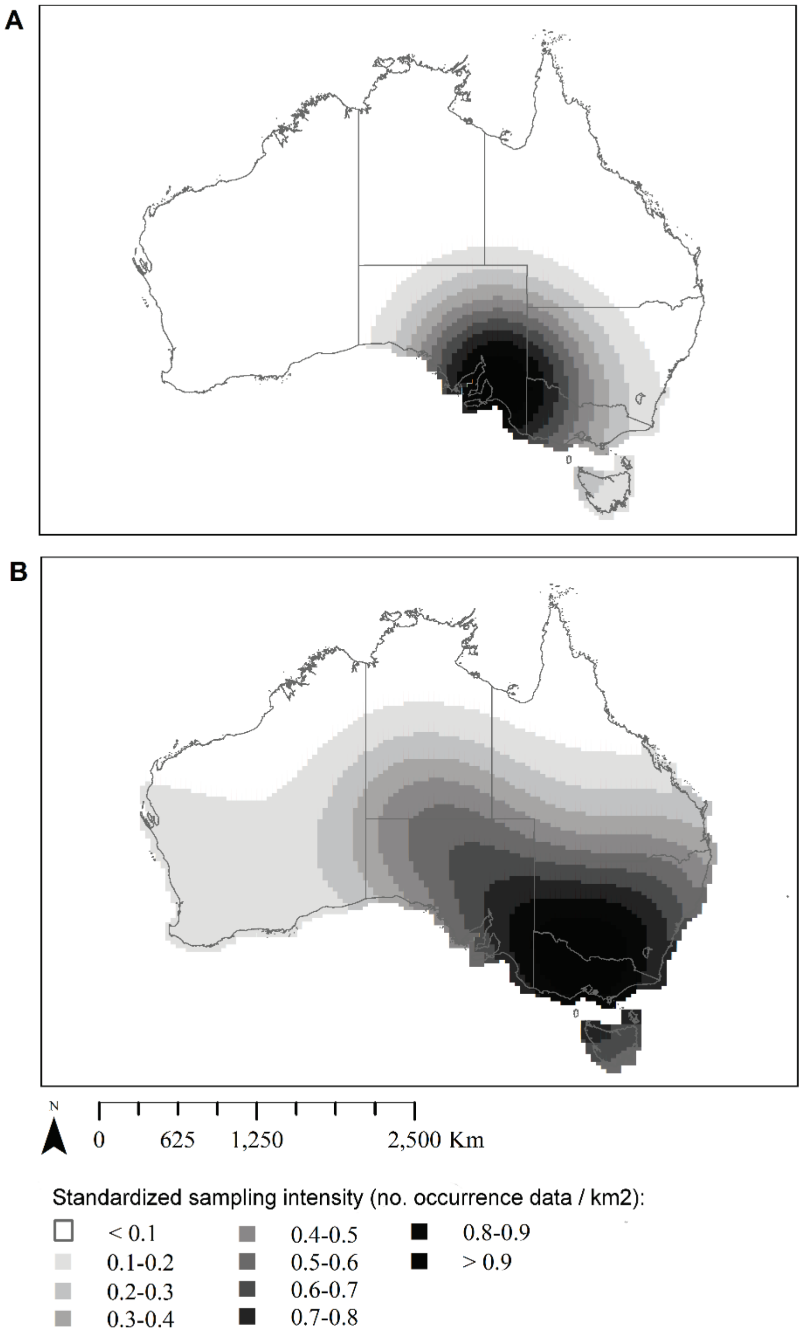

Figure 9: Estimated geographical bias in the sampling effort based on the standardised density of occurrence data collated in the *Australian National Rabbit Database*. (A) Sampling intensity calculated on all occurrence points. (B) Sampling intensity calculated using only one occurrence point per 10 km^2 grid.

e. Quality assessment

A dataset may suffer from several weaknesses that can decrease the quality of the products derived from it (Hortal *et al.* 2007, Robertson *et al.* 2010, and

Pipino *et al.* 2002). Therefore, we developed a systematic and comparative scheme to evaluate and rank the data quality compiled in the database based on four types of potential issues: (1) data accuracy, (2) data completeness, (3) temporal coverage, and (4) data consistency (Figure 10, see Supplementary Information SI8).

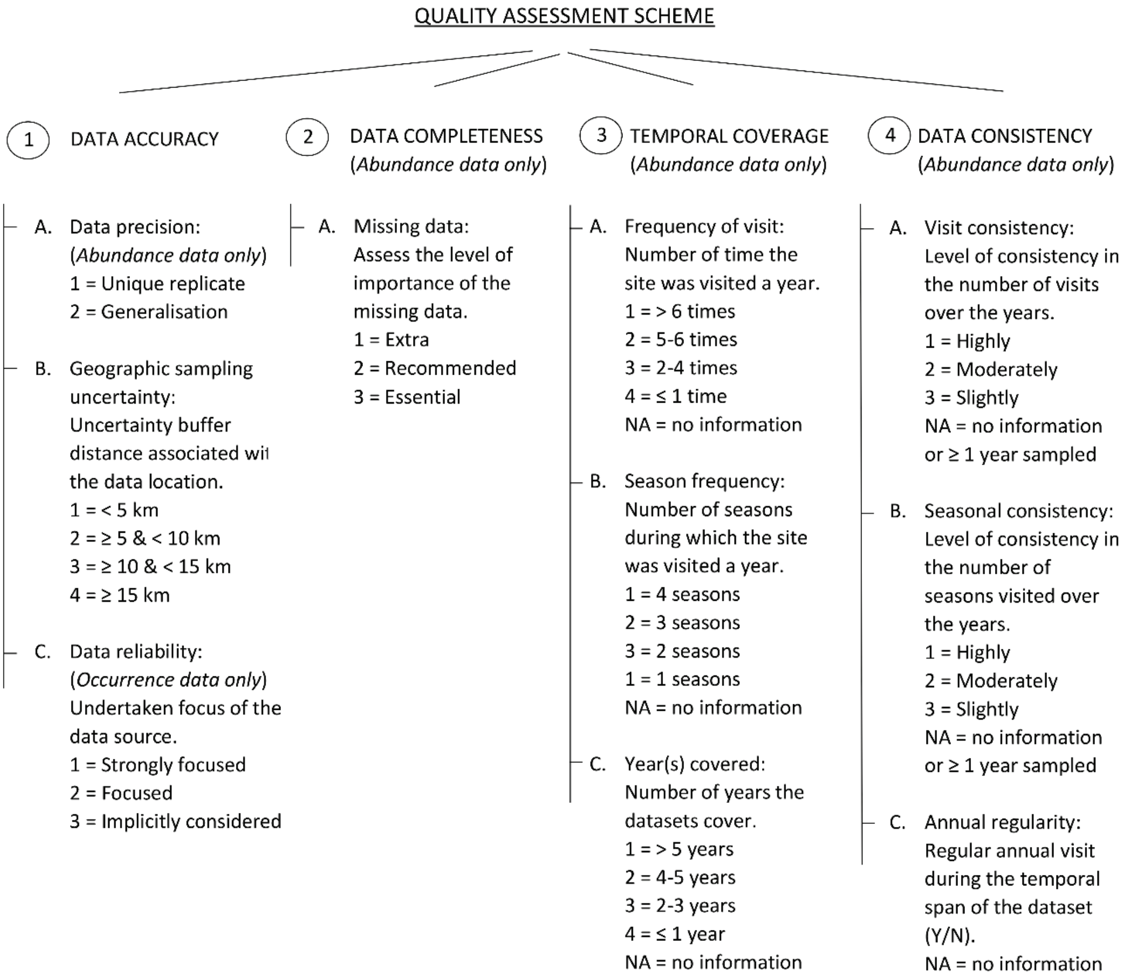

Figure 10: Quality assessment scheme used to rank the data. We assessed four main data weaknesses (i.e. data accuracy, completeness, temporal coverage, and consistency) using a set of sub-criteria (listed with the capitalised letters).

<u>Data accuracy</u> indicates the accuracy of the data and their reliability based on three sub-criteria: (*i*) the precision of the count data, (*ii*) the geographic location sampling uncertainties, and (*iii*) the reliability of the data. Count data precision (abundance data only) was estimated based on the presence or absence of replicates data. Data with more than one replicate were classified in category

‘one’, while generalised data (i.e. average of replicates data) was classified in category ‘two’. Similarly, the geographic sampling uncertainty was determined by the information field ‘resolution uncertainty’ which has four levels: < 5 five km, ≥ 5 km and < 10 km, ≥ 10 km and < 15 km, and ≥ 15 km (category from one to four respectively). These distances are radius lengths around the data location corresponding to an ‘uncertainty buffer distance’ associated with the current location of the data. These levels were assigned systematically to every data entry based on information provided on the location where the data was sampled (i.e. from accurate geographical coordinates to estimated anecdotal location). Lastly, the data reliability (occurrence data only) classifies the data into three categories: implicitly considered, focused, and strongly focused. These categories are based on the National Land & Water Resources Audit and Invasive Animals Cooperative Research Centre (2008). The first category (strongly focused data), follows standard field-sampling protocols: surveys, systematic sampling, or formal assessment including spotlight observations, disease monitoring site, and shooting activities. The second category (focused data), includes indirect data from experts collected during control activities, including warren ripping, collection of presence signs, and the release of biological agents (i.e. Myxoma virus and RHDV). The last category (implicitly considered data), results from second resource information such as anecdotal sightings, incidental reports, and less reliable expert knowledge or survey data obtained from citizen science projects.

Data completeness is the extent to which a data are sufficient to complete a given task (Pipino *et al*. 2002) and is provided only for the abundance data. The abundance data have 64 information fields, but four are mandatory (see section Class II.B.2.A, i.e. geolocation of the data including the latitude and longitude, year of the sampling process, and the estimated number of observations per kilometres). If the information associated with the data only fills these fields, the data are classified as category ‘three’ (i.e. essential information). Six extra information fields are highly recommended to evaluate other data quality criteria (such as the accuracy of the geographic location), to derive precise weather, climatic, and environmental variables extractions (category two – recommended information). These fields include the data type, the site section name, and the month and day corresponding to the sampling date, the total number of replicates used to generate the abundance estimations, and the transect length. The remaining fields are 'optional' and are only found in dataset with extra set of information (category one).

<u>Temporal coverage</u> is important for time series analyses. Therefore, for the abundance dataset, we provide (*i*) the number of visits per year, (*ii*) the number of seasons for which the site was visited per year, and (*iii*) the number of years of sampling. The number of visits per year is calculated as the mode value over all years, or the median if a different number of visits was made every year at the site. We classified the number of visits into four categories: > 6 times, between 5-6 times, between 2-4 times, and ≤ 1 time only (category one to four, respectively). We also calculated the number of seasons the fieldwork was carried out (by using mode value of the number of seasons per year) and classified the results into five categories: from one to four seasons and 'NA' when monthly information was missing. Finally, the number of years of available data were classified as follow: > 5 years, between 4-5 years, between 2-3 years, and ≤ 1 year (category one to four respectively) and extra categories 'NA' indicate missing information.

<u>Consistency</u> quantifies the extent to which the data point presents any inconsistencies at the level of sampling effort and major temporal breaks in the data collection. It does this by assessing, for every dataset, the regularity of: (*i*) the sampling visits over the time period of data collection (called hereafter 'visit consistency'), (*ii*) the number of seasons during which the site was visited each year (called hereafter 'seasonal consistency'), and (*iii*) yearly visit for a given site (called hereafter 'annual regularity'). The 'consistency' is calculated as the coefficient of variation for each site standardised by the number of sampled years. We classified the results of 'visit consistency' into three categories: highly consistent (< 8), moderately consistent (between 8 and 11), and slightly consistent (≥ 11). Values to classify the 'seasonal consistency' are: < 0.075 (highly consistent), between 0.075 and 0.09 (moderately consistent), and ≥ 0.09 (slightly consistent). A 'NA' indicates that data were sampled during only one year (i.e. resulting in no variation), or if monthly information is not available to

calculate the seasonal consistency. Regular annual visits for a given site is noted as 'Y', otherwise, an 'N' indicates the presence of a gap year(s), and a 'NA' specified missing information to classify the data.

f. Data limitations:

The *Australian National Rabbit Database* includes many different types of data (e.g. expert or citizen providers). As a result, the quality and reliability of the information provided by each dataset may vary. To account for these limitations in our database, we included a measure of the data quality (see section Class II.B.3.E). We, however, recommend that users check previous land management activities at each site, such as land clearing and agriculture, and prevalence of diseases that could have biased data collection.

Furthermore, a variety of factors can influence the detection probability of rabbits. The behaviour of rabbits can, for instance, change due to variable weather conditions (Twigg *et al.* 1998, Ballinger & Morgan, 2002). Likewise, vegetation cover along transects can also influence detection probability. If these potential sources of imperfect detection are not taken into account when estimating relative abundance, the results obtained can be biased (Royle & Nichols, 2003). For this reason, we do not provide a density estimate per grid cell as part of this database. We recommend users to employ statistical approaches that allow for sampling and temporal biases in abundance count data to be identified and explicitly accounted for in their analyses of the database (e.g. Joseph *et al.* 2009, Wells *et al.* 2016a).

CLASS III. DATA SET STATUS AND ACCESSIBILITY

A. Status:

1. Latest update:

February 2019.

2. Latest archive date:

3. February 2019.

4. **Metadata status:**

   Last updated February 2019, submitted version.

5. **Data verification:**

   We followed the methodological framework presented in the section Class II.B.3.b to maintain a standardised format of consistent data and to reduce potential sources of errors. The quality of the datasets was further assessed by exploring the datasets (see section Class I.C.E), as well as conducting preliminary analyses which we provide as step-by-step tutorials in HTML format (see Supplementary Information SI9 and SI10).

B. **Accessibility:**

1. **Storage location and medium:**

   The dataset can be accessed as supporting information to this Data Paper publication in *Ecology:* see DataS1.zip. The dataset and additional information (such as step-by-step tutorials in HTML format) are also available within DataS1.zip (Supplementary Information S9 and S10).

2. **Contact persons:**

Emilie Roy-Dufresne

The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

roydufresne.emilie@gmail.com

Dr Damien Fordham

The Environment Institute and School of Biological Sciences, University of Adelaide, SA 5005, Australia.

damien.fordham@adelaide.edu.au

3. **Copyright restrictions:**

   None.

4. **Proprietary restrictions:**

   a. **Release data:**

      None.

b. **Citation:**

Please cite this data paper when the data are used in publications or teaching events.

c. **Disclaimers:**

None.

5. **Costs:**

None.

CLASS IV. DATA STRUCTURAL DESCRIPTORS

A. **Data set file:**

We divided the dataset into two main features: occurrence and abundance data. Occurrence data are observations of one or more rabbits, while the abundance data are counts of the rabbits made along transects at night with a spotlight. For both, the information is structured so that each row contains a single and unique record.

1. **Identity:**
    a. AustralianNationalRabbitDatabase_OccurrenceData.csv
    b. AustralianNationalRabbitDatabase_AbundanceData.csv

2. **Size:**
    a. 36 columns; 689,265 rows; 168 MB
    b. 64 columns; 51,241 rows; 16.6 MB

3. **Format and storage mode:**
    a. comma-separated values (.csv)
    b. comma-separated values (.csv)

4. **Header information:**

    See column descriptions in section Class IV.B.1.

5. **Alphanumeric attributes:**

Mixed.

B. **Variable information:**

Table 6. Information on variables for the occurrence data.

Column	Title	Format	Description
1	Occurrence_ID	Int	Unique ID # for each occurrence data
2	AustraliaUTMGrid10Km_ID [1,2]	Text	ID of the UTM grid cell in where the observation was made
3	AustraliaUTMGrid10Km_Lat [2]	Num	Latitude associated with the UTM grid (WGS84)
4	AustraliaUTMGrid10Km_Long [2]	Num	Longitude associated with the UTM grid (WGS84)
5	ResolutionUncertainty[3]	Text	Geographical location uncertainty value associated with the data
6	TotalNbOccurrencesPer UTMGrid [2]	Int	Number of occurrence points falling within each UTM grids
7	PtsIntensityEstimate_AllData PerUTMGrid[2]	Num	Number (standardised between 0 and 1) of random observations that are expected to be collected in each grid based on the current spatial pattern observed in the occurrence data – including all occurrence data (see section Class II.B.3.d)
8	PtsIntensityEstimate_ OneDataPerUTMGrid[2]	Num	Number (standardised between 0 and 1) of random observations

			that are expected to be collected in each grid based on the current spatial pattern observed in the occurrence data – including only one occurrence data per grid cell (see section Class II.B.3.d)
9	AustraliaState [2,4]	Factor	Australian states where the observation was made
10	DataType [5]	Factor	Method used to obtain the data (i.e. observations from experts or citizens providers, observations from spotlight or presence signs monitoring researches, observations from warren, shooting, or RHDV1 management actions, and observations associated to Spanish Flea's release)
11	DataSourceEntry_ID	Text	Unique ID assigned to each entry within a source of data
12	DataSourceName	Text	Name of the source of data
13	DataSourceSiteSectionName	Text	Sub-section of the study site as defined by the data provider
14	Year	Int	Year at which the species was observed
15	Month	Int	Month at which the species was observed
16	Day	Int	Day at which the species was observed
17	Season[6]	Factor	Season corresponding to the sampling month following the Australian calendar
18	QA_Accuracy[2]	Num	Quality assessment of the data accuracy (see section Class II.B.3.e)

19	A_Prec_Avg30Yr	Num	30-yrs mean (1976-2005) annual precipitation (mm)
20	A_Psea_Avg30Yr	Num	30-yrs mean (1976-2005) annual precipitation seasonality (mm)
21	A_TAvg_Avg30Yr	Num	30-yrs mean (1976-2005) annual average temperature (°C)
22	A_TMax_Avg30Yr	Num	30-yrs mean (1976-2005) maximum temperature of warmest month (°C)
23	A_TMin_Avg30Yr	Num	30-yrs mean (1976-2005) minimum temperature of coldest month (°C)
24	A_TSea_Avg30Yr	Num	30-yrs mean (1976-2005) annual temperature seasonality (°C)
25	A_TWet_Avg30Yr	Num	30-yrs mean (1976-2005) of mean temperature of wettest quarter (°C)
26	A_TWrm_Avg30Yr	Num	30-yrs mean (1976-2005) of mean temperature of warmest quarter (°C)
27	A_Prec_AvgAutumn30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in autumn (mm)
28	A_Prec_AvgSummer30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in summer (mm)
29	A_Prec_AvgSpring30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in spring (mm)

30	A_Prec_AvgWinter30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in winter (mm)
31	VegeType [7]	Factor	Major vegetation types for the Australian land (see section Class V.D)
32	DistPermWater	Num	Weighted Euclidean distance to permanent water features (km) (see section Class V.D)
33	DistAgriLand	Num	Weighted Euclidean distance to agricultural land margins (km) (see section Class V.D)
34	PercSoilClay	Num	Percentage of clay (%)(see section Class V.D)
35	MinDayLength	Num	Minimum day length across Australia (hours) (see section Class V.D)
36	VarDayLength	Num	Annual variance in day length across Australia (hours) (see section Class V.D)

[1] UTM grid numbers consist of the official UTM number and letter as defined by the projection coordinates followed by a unique number that represents the sub-section of the UTM grid ordered from West to East and North to South (maximum number: 4800).

[2] R script used to process the information provided on (see Supplementary information SI5).

[3] Factor level: less than 5 km, between 5 km and 10 km, between 10 km and 15 km, more than 15 km.

[4] Factor level: WA, NT, SA, QLD, NSW, ACT, VIC, and TAS.

[5] Factor level: Expert observations, citizen observations, spotlight transects, warren locations, presence sign quadrats, Spanish fleas release sites, RHV1 spread survey sites, and shooting sites.

[6] Factor level: 1 (Summer : Jan, Feb, Mar), 2 (Autumn: Apr, May, Jun), 3 (Winter: Jul, Aug, Sep), 4 (Spring: Oct, Nov, Dec)

[7] Factor level: 13 classes define the vegetation structure. The list is provided in section Class V.D.

Table 7. Information on variables for the abundance data.

Column	Title	Format	Description
1	SpotData_ID	Num	Unique ID # for each spotlight data
2	AustraliaUTMGrid10Km_ID [1,2]	Text	ID of the UTM grid in which the observation was made
3	AustraliaUTMGrid10Km_Long [2]	Num	Longitude associated with the UTM grid (WGS84)
4	AustraliaUTMGrid10Km_Lat [2]	Num	Latitude associated with the UTM grid (WGS84)
5	ResolutionUncertainty[3]	Factor	Geographical location uncertainty value associated with the data
6	TotalNbTransectsPerUTMGrid [2]	Int	Total number of transects located in the UTM grid
7	AustraliaState [2,4]	Factor	Australian state where the data was collected
8	StudyRegions_ID	Num	ID # for the study regions (Geographic area defined as site by the data provider)
9	DataSourceEntry_ID	Text	Unique ID assigned to each entry within a source of data
10	DataSourceName	Text	Name of the source of data

11	DataSourceSiteSectionName	Text	Sub-section of the study site as defined by the data provider
12	VisitNb	Int	Number associated with the act of visiting the region of the study. For each region of study, the number increases chronological providing the total number of time the region was visited. The visit numbers were processed in the database as defined by the data provider.
13	TotalNbReplicates[2]	Int	Total number of replicates collected and provided per visit.
14	Year	Int	Year the datum was collected
15	Month	Int	Month the datum was collected
16	Day	Int	Day the datum was collected
17	Season[5]	Factor	Season corresponding to the sampling month following the Australian calendar
18	Transect_ID	Factor	Unique ID given to each transect as defined by the data provider
19	TransectWidth	Num	Distance illuminated by the spotlight on each side of the transect in meter
20	TransectLength	Num	Length of the transect in kilometre
21	RabbitsPerKm	Num	Number of rabbits observed per kilometre
22	UncertaintyManagementNotes	Text	Any notes, as defined by the data provider, which could results in additional uncertainties and impact the reported counts (e.g. disease outbreak, previous or current management activities)
23	QA_Accuracy[2]	Num	Quality assessment of the data accuracy (see section Class II.B.3.e)

24	QA_Completeness[2]	Num	Quality assessment of the data completeness (see section Class II.B.3.e)
25	QA_TemporalCoverage[2]	Num	Quality assessment of the temporal data coverage (see section Class II.B.3.e)
26	QA_Consistency[2]	Num	Quality assessment of the data consistency (see section Class II.B.3.e)
27	D_Prec	Num	Average daily rainfall for the same day of the sampling date (mm)
28	D_Prec_30DaysLag	Num	Mean rainfall for 30 days before the sampling date (mm)
29	M_Prec	Num	Average monthly rainfall for the same month of the sampling date (mm)
30	M_Prec_12PreMonths	Num	Twelve months rainfall average for a year time-range period before the data sampling date (mm) (i.e. one-year time lag)
31	M_Prec_24Months	Num	Twelve months rainfall average for a two years time-range period before the data sampling date (mm) (i.e. two-year time lag)
32	M_Prec_2PreSeasons	Num	Mean rainfall of the two previous seasons (mm) Seasons are estimated based on the Australian calendar
33	A_Prec	Num	Average annual rainfall for the same year of the sampling date (mm)
34	A_Prec_1YrLag	Num	Average annual rainfall for the year before the sampling date (mm) (i.e. one-year time lag)

35	A_Prec_2YrLag	Num	Average annual rainfall for two years before the sampling date (mm) (i.e. two-year time lag)
36	A_Psea	Num	Seasonal average rainfall for the current year of the data sampling (mm)
37	A_Psea_1YrLag	Num	Seasonal average rainfall for the year before the sampling date (mm) (i.e. one-year time lag)
38	A_Psea_2YrLag	Num	Seasonal average rainfall for two years before the sampling date (mm) (i.e. two-year time lag)
39	D_Tmax	Num	Average daily maximum temperature for the same day of the sampling date (°C)
40	D_Tmin	Num	Average daily minimum temperature for the same day of the sampling date (°C)
41	M_TAvg_PreSummer	Num	Average monthly temperature for the previous summer season (°C) Seasons are estimated based on the Australian calendar
42	M_TAvg_PreWinter	Num	Average monthly temperature for the previous winter season (°C) Seasons are estimated based on the Australian calendar
43	A_AvgTMaxWarmestMonth	Num	Average monthly maximum temperature for the warmest month for the same year the sampling date (°C)
44	A_AvgTMaxWarmestMonth_1YrLag	Num	Average monthly maximum temperature for the warmest month for the previous year the sampling

			date (°C) (i.e. one-year time lag)
45	A_AvgTMinColdestMonth	Num	Average monthly minimum temperature for the coldest month for the same year the sampling date (°C)
46	A_AvgTMinColdestMonth_ 1YrLag	Num	Average monthly minimum temperature for the coldest month for the previous year the sampling date (°C) (i.e. one-year time lag)
47	A_Prec_Avg30Yr	Num	30-yrs mean (1976-2005) of annual precipitation (°C)
48	A_Psea_Avg30Yr	Num	30-yrs mean (1976-2005) of annual precipitation seasonality (°C)
49	A_TAvg_Avg30Yr	Num	30-yrs mean (1976-2005) of annual mean temperature (°C)
50	A_TMax_Avg30Yr	Num	30-yrs mean (1976-2005) maximum temperature of warmest month (°C)
51	A_TMin_Avg30Yr	Num	30-yrs mean (1976-2005) minimum temperature of coldest month (°C)
52	A_TSea_Avg30Yr	Num	30-yrs mean (1976-2005) of annual temperature seasonality (°C)
53	A_TWet_Avg30Yr	Num	30-yrs mean (1976-2005) of mean temperature of wettest quarter (°C)
54	A_TWrm_Avg30Yr	Num	30-yrs mean (1976-2005) of mean temperature of warmest quarter (°C)
55	A_Prec_AvgAutumn30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in autumn (mm)

56	A_Prec_AvgSummer30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in summer (mm)
57	A_Prec_AvgSpring30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in spring (mm)
58	A_Prec_AvgWinter30Yr	Num	30-yrs mean (1976-2005) of annual precipitation in winter (mm)
59	VegeType [6]	Factor	Major vegetation types for the Australian land (13 Classes) (see section Class V.D)
60	DistPermWater	Num	Euclidean weighted distance to permanent water features (km) (see section Class V.D)
61	DistAgriLand	Num	Euclidean distance to agricultural land margins (km) (see section Class V.D)
62	PercSoilClay	Num	Percentage of clay (%)(see section Class V.D)
63	MinDayLength	Num	Minimum day length across Australia (hours) (see section Class V.D)
64	VarDayLength	Num	Annual variance in day length across Australia (hours) (see section Class V.D)

[1] UTM grid numbers consist of the official UTM number and letter as defined by the projection coordinates followed by a unique number representing the sub-section of the UTM grid ordered from West to East and North to South (maximum number: 4800).

[2] R script used to process the information provided on (see Supplementary Information SI6).

[3] Factor level: less than 5 km, between 5 km and 10 km, between 10 km and 15 km, more than 15 km.

[4] Factor level: WA, NT, SA, QLD, NSW, ACT, VIC, TAS

[5] Factor level: 1 (Summer: Jan, Feb, Mar), 2 (Autumn: Apr, May, Jun), 3 (Winter: Jul, Aug, Sep), 4 (Spring: Oct, Nov, Dec)

[6] Factor level: 13 classes define the vegetation structure. The list is provided in section Class V.D.

C. Data anomalies:

We addressed each data error if the information was available, otherwise we indicated the data field as missing or unavailable using 'NA'.

CLASS V. SUPPLEMENTAL DESCRIPTORS

A. Data acquisition:

1. Original data format or acquisition methods:

We extracted the primary data from multiple sources of mediums: field notes, excel spreadsheets, access tables, floppy disks, and geographic information software (e.g., ArcGIS, ESRI 2016; and MapInfo Professional, Pitney Bowes Software 2017).

2. Location of completed data forms:

The completed data forms will be provided to the Centre for Invasive Species Solutions (https://invasives.com.au/).

3. Data entry verification procedures:

We followed the methodological framework presented in the section Class II.A.3.b to maintain a standardised format of consistent data and to reduce potential sources of errors. The quality of the datasets was further assessed by exploring the general structure of the datasets (see section Class I.C.E), as well as performing some pre-analysis which we provided as step-by-step tutorials in HTML format (see Supplementary Information SI9 and SI10).

C. Related materials:

To demonstrate how to extract the information from the database and to explore aspects of the data, we created two step-by-step tutorials in HTML format (see Supplementary Information SI9 and SI10).

D. Computer programs and data-processing algorithms:

We processed and included six environmental variables when building the database. Table 8 lists those variables, their descriptions, and the methodological process for formatting them for the database. Variables that did not follow a normal distribution were log-transformed (noted as log-transformation in Table 8) to reduce any potential nonlinearity of their responses and to improve model performance (Austin 2002).

Table 8: Environmental variables included in the database.

Variables	Description	Processing tasks
VegeType[1]	Major vegetation types for the Australian land (13 Classes)	We reclassified the major vegetation groups of the National Vegetation Information System (NVIS; version 4.1) into 13 categories. See the re-classification list below.
DistPermWater[2,3]	Weighted Euclidean distance to permanent water features	We calculated the Euclidean distance in kilometre using ArcGIS 10.3.1 to any permanent water features as provided by the CSIRO and surface hydrology points (i.e. farm dam water, native well, water hole, and water tank) provided by Geoscience Australia. Log transformation.
DistAgriLand[4]	Weighted Euclidean distance to agricultural land margins	We sub-sampled the crop types (i.e. cropping, grazing irrigated modified pastures, irrigated cropping, perennial and seasonal horticulture, and irrigated perennial and seasonal horticulture, intensive horticulture, and intensive animal husbandry) within the land use dataset and we calculated the pairwise Euclidean distance in kilometre between each of them using ArcGIS 10.3.1. Log transformation.
PercSoilClay[5]	Percentage of clay	--
MinDayLength	Minimum day length across Australia	We calculated the day length in hours across Australia using the geosphere package in R, and we extracted the lowest value over a year for every site locations.

VarDayLength	Annual variance in day length across Australia	We calculated the day length in hours across Australia using the geosphere package in R, and we calculated the yearly variance for every site locations.

[1] Data obtained from the Australian Major and Sub-Vegetation Groups dataset from the Environment Department of the Australian Government (http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products)

[2] Data obtained from CSIRO from the Atlas of Living Australia (http://www.ala.org.au/)

[3] Data obtained from Geoscience Australia (http://www.ga.gov.au/)

[4] Data from the Department of Agriculture of the Australian Government (http://data.daff.gov.au/anrdl/metadata_files/pa_luav4g9abl07811a00.xml)

[5] Data obtained from the Australian Soil Resource Information System hosted by CSIRO (http://www.asris.csiro.au/themes/Atlas.html)

We re-classified the VegeType variable into 13 categories to reduce the number of factors included in the analysis (Figure 11 and 12) as follow: (*i*) rainforest and vine thicket as in the initial variable, (*ii*) eucalyptus forest (i.e. open forests, low open forests, open woodlands, and tropical woodlands/grasslands), (*iii*) eucalyptus woodland (see source initial variable), and also mallee woodlands, mallee open woodlands, and sparse mallee shrublands, (*iv*) woodlands category that includes acacia forests, open acacia woodlands, callitris forests, casuarina forests, other forests and woodlands, and other open woodlands, (*v*) seasonal inundated swamps, salt marshes and mangroves, melaleuca forests, other grasslands, herblands, sedgelands, rushlands, and estuaries, (*vi*) low-closed forests and tall closed shrublands (including Acacia, Melaleuca and Banksia), (*vii*) shrublands (i.e., acacia, heathlands, and other shrublands), (*viii*) tussock grasslands, (*ix*) hummock grasslands, (*x*) chenopod shrublands, samphire shrublands, and forblands under the saltbushes category, (*xi*) inland aquatic, fresh water, salt lakes, estuaries, and lagoons, (*xii*) cleared vegetation, non-native vegetation, buildings, rocks, claypan, mudflat, and naturally bare areas, and (*xiii*) unclassified and unknown features.

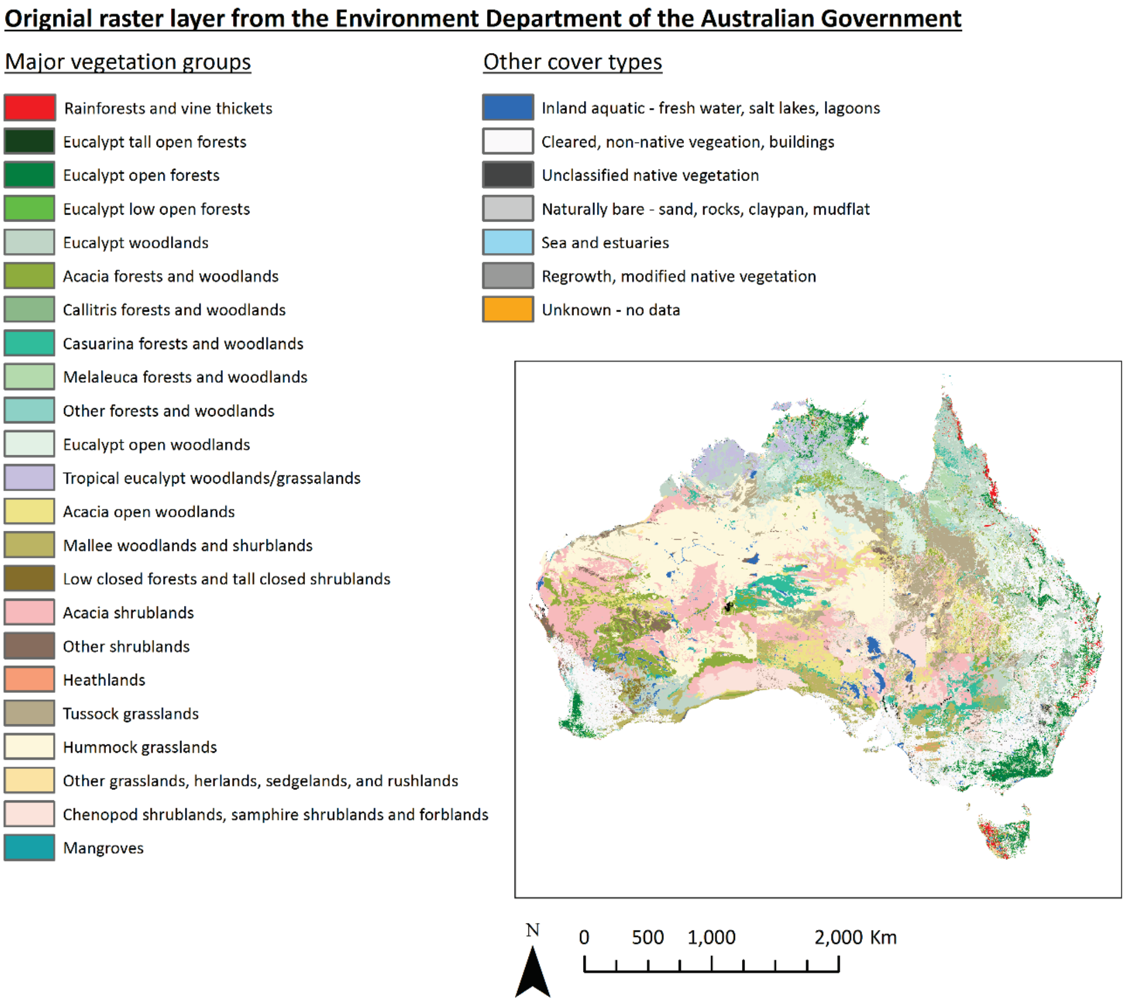

Figure 11. Distribution of the major and sub-grouping vegetation types in Australia. According to the Environment Department of the Australian Government classification scheme (http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products).

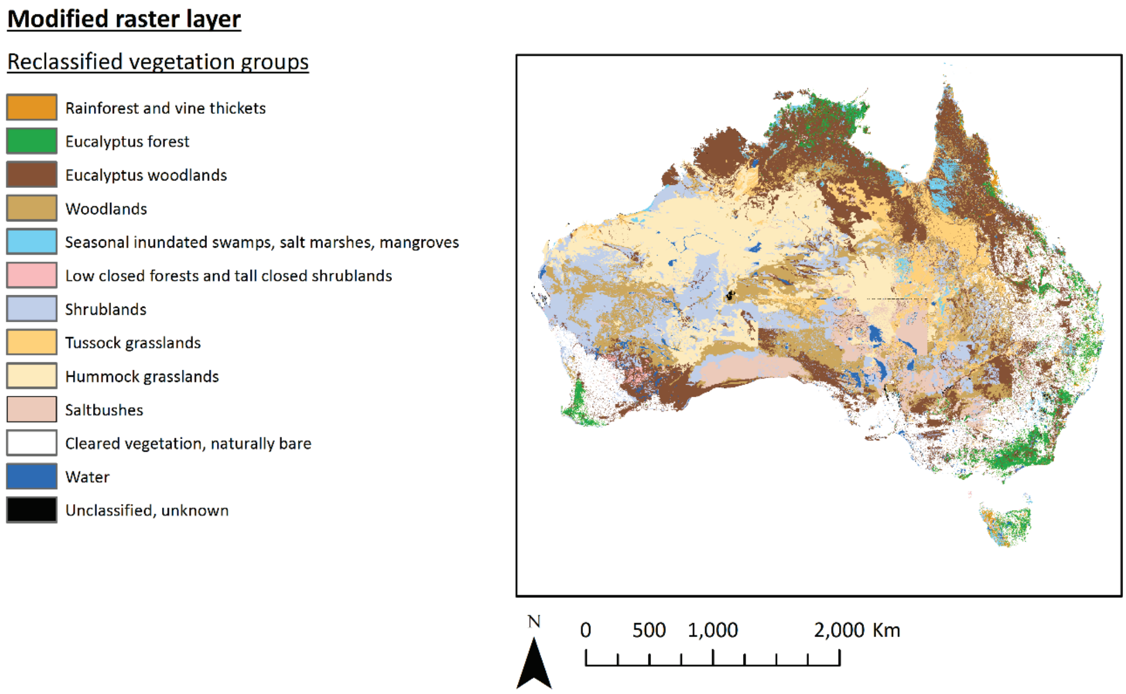

Figure 12. Reclassification of the original 'Vegetype' raster layer into the 13 categories of the Australian National Dataset. Data obtained from the Australian Major and Sub-Vegetation Groups dataset from the Environment Department of the Australian Government classification scheme (http://www.environment.gov.au/land/native-vegetation/national-vegetation-information-system/data-products).

We developed R scripts program (R Core Team 2016) to process the information from the database; see the list below (in Supplementary Information):

1. SI5_NewAttributes_OccurrenceData.R

   Prepare the new attributes for the occurrence data.

2. SI6_NewAttributes_AbundanceData.R

   Prepare the new attributes for the abundance data.

3. SI7_KSmoothIntensity.R

   Estimate the geographical bias associated with the spatial pattern of the occurrence data.

4. SI8_QualityAssessment.R

   Rank the data quality according to our quality assessment scheme: (1) data accuracy, (2) data completeness, (3) temporal coverage, and (4) data consistency.

We also provided two steps-by-steps tutorials (R format language) to demonstrate how to process the occurrence and abundance data (e.g. extracting required information for a particular site base on geographical coordinates) and to achieve different analysis (e.g. generating summary maps and time series plots; in Supplementary Information):

1. SI9_DataTutorial_OccurrenceData.html
   Steps-by-steps tutorial for the occurrence data.
2. SI10_DataTutorial_AbundanceData.html
   Steps-by-steps tutorial for the abundance data.

E. Publication and results:

A list of references associated with each abundance dataset providing details on the original aims and techniques employed during the sampling process is provided in the file 'SI4_DataInfo_AbundanceData.pdf' (in Supplementary Information).

Other publications which analysed the data collated within the database include:

Lurgi, M., K. Wells, M. Kennedy, S. Campbell, D. A. Fordham. 2016. A landscape approach to invasive species management. PlosOne 11: e0160417. doi.org/10.1371/journal.pone.0160417

Lurgi, M., E. G. Ritchie, D. A. Fordham. 2018. Eradicating abundant invasive prey could cause unexpected and varied biodiversity outcomes: The importance of multispecies interactions. Journal of Applied Ecology 55: 2396-2407. doi.org/10.1111/1365-2664.13188.

F. History of data set usage:

1. **Data request history:**
   None.

2. **Data set update history:**
   None.

3. **Review history:**

   None.

4. **Questions and comments from secondary users:**

   None.

ACKNOWLEDGEMENTS

This project and the authors were supported by the Australian Research Council (LP110200805). The authors wish to acknowledge the global participation of the various data owners who gave their time, advice, access to their data and data systems generously. These owners include governmental stakeholders, scientific experts, BHP, monitoring contractors, station owners, and the broader community.

LITERATURE CITED

Anderson, A. 2009. The rat and the octopus: initial human colonisation and the prehistoric introduction of domestic animals to Remote Oceania. Biological Invasions 11:1503–1519.

Ashok, K., Z. Guan, and T. Yamagata. 2003. Geophysical Research Letters 30: 1821.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101-118.

Baddeley, A., E. Rubak, and R. Turner. 2015. Spatial point patterns: methodology and applications with R. Chapman and Hall/CR Press, Long

Ballinger, A., and D. G. Morgan. 2002. Validating two methods for monitoring population size of the European rabbit (*Oryctolagus cuniculus*). Wildlife Research 29, 431.

Bird, P., G. Mutze, D. Peacock, and S. Jennings. 2012. Damage caused by low-density exotic herbivore populations: the impact of introduced European rabbits on marsupial herbivores and *Allocasuarina* and *Bursaria* seedling survival in Australian coastal shrubland. Biological Invasions 14:743-755.

Blewett, R. S. 2012. Shaping a nation: a geology of Australia. Geoscience Australia and ANU E-Press, Canberra.

Bowen, Z., and J. Read. 1998. Population and demographic pattern of rabbits (*Oryctolagus cuniculus*) at Roxby Downs in arid South Australia and the influence of rabbit haemorrhagic disease. CSIRO Wildlife Research 25: 655-662.

Bowler, J. M. 1976. Aridity in Australia, age, origins and expression in Aeolian landforms and sediments. Earth-Science Reviews 12: 279-310.

Cooke, B. D. 1970. Factors which influence the number of rabbits, *Oryctolagus cuniculus* (L.) in natural populations. Master Thesis, University of Adelaide.

Cooke, B. D. 2018. Long-term monitoring of disease impact: rabbit haemorrhagic disease as a biological control case study. Veterinary Record 182:571-572.

Cooke, B., P. Chudleigh, S. Simpson, and G. Saunders. 2013. The economic benefits of the biological control of rabbits in Australia, 1950–2011. Australian Economic History Review 53:91-107.

Dennis, R. L. H., T. G. Shreeve, N. J. B. Isaac, D. B. Roy, P. B. Hardy, R. Fox, and J. Asher. 2006. The effects of visual apparency on bias in butterfly recording and monitoring. Biological Conservation 128:486-492.

England, M., C. C. Ummenhofer, and A. Santoso. 2005. Journal of Climate 19: 1948-1969.

ESRI. 2016. ArcGIS Desktop. Environmental Systems Research Institute, Redlands, California, US.

Fenner, F., and B. Fantini. 1999. Biological control of vertebrate pests: the history of myxomatosis, and experiment in evolution. CABI Publishing, Wallingford, UK.

Fordham, D. A., R. G. Sinclair, D. E. Peacock, G. J. Mutze, J. Kovaliski, P. Cassey, L. Capucci, and B. W. Brook. 2012. European rabbit survival and recruitment are linked to epidemiological and environmental conditions in their exotic range. Austral Ecology 37: 945-957.

Forsyth, D. M., M. P. Scroggie, A. D. Arthur, M. Lindeman, D. S. L. Ramsey, S. R. McPhee, T. Bloomfield, and I. G. Stuart. 2015. Density-dependent effects of a widespread invasive herbivore on tree survival and biomass during reforestation. Ecosphere 6(4):71.

Fourcade, Y., J. O. Engler, D. Rödder, and J. Secondi. 2014. Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. PlosOne 9:e97122.

Gilbert, N., K. Myers, B. D. Cooke, J. D. Dunsmore, P. J. Fullagar, J. A. Gibb, D. R. King, I. Parer, S. H. Wheeler, and D. H. Wood. 1987. Comparative dynamics of Australasian rabbit populations. Australian Wildlife Research 14: 491-503.

Gong, W., T. Sinden, M. L. Braysher, R. Jones, and N. S. Wales. 2009. The economic impacts of vertebrate pests in Australia. Invasive Animals Cooperative Research Centre, Canberra. P. 60.

Henzell, R. P., B. D. Cooke, and G. J. Mutze. 2008. The future biological control of pest populations of European rabbits, Oryctolagus cuniculus. Widlife Research 35: 633-650.

Henzell, R. P., R. B. Cunningham, and H. M. Neave. 2002. Factors affecting the survival of Australian wild rabbits exposed to rabbit haemorrhagic disease. Wildlife Research 29: 523-542.

Hijmans, J., M. Schreuder, J. De la Cruz, and L. Guarino. 1999. Using GIS to check co-ordinates of genebank accessions. Genetic Resources and Crop Evolution 46:291–296.

Hortal, J., J. M. Lobo, and A. Jiménez-Valverde. 2007. Limitations of Biodiversity Databases: Case Study on Seed-Plant Diversity in Tenerife, Canary Islands. Conservation Biology 21:853–863.

Hutchinson, M. F., J. Kesteven, and T. Xu. 2014. ANUClimate 1.0, 0.01 degree, Australian Coverage, 1970-2014. Australian National University, Canberra, Australia. Obtained from http://dap.nci.org.au. Accessed June 28th 2018.

Hutchinson, M. F., D. W. Mckenney, K. Lawrence, J. Pedlar, R. Hopkinson, E. Milewska, and P. Papadopol. 2009. Development and testing of Canada-wide interpolated spatial models of daily minimum/maximum temperature and precipitation for 1961-2003. Journal of Applied Meteorology and Climatology 48: 725–741.

Hutchinson, M.F., and T. Xu. 2013. ANUSPLIN version 4.4 User Guide. Fenner School of Environment and Society, Australian National University, Canberra. Obtained from http://fennerschool.anu.edu.au/files/anusplin44.pdf

Joseph, L. N., C. Elkin, T. G. Martin and H. P. Possingham. 2009. Modeling abundance using N-mixture models: the importance of considering ecological mechanisms. Ecological Applications 19(3): 631-642.

Lurgi, M., K. Wells, M. Kennedy, S. Campbell, and D. A. Fordham. 2016. A landscape approach to invasive species management. PlosOne 11: e0160417.

Lurgi, M., E. G. Ritchie, and D. A. Fordham. 2018. Eradicating abundant invasive prey could cause unexpected and varied biodiversity outcomes: The importance of multispecies interactions. Journal of Applied Ecology, doi.org/10.1111/1365-2664.13188.

McKenzie, N. J., and P. J. Ryan. 1999. Spatial prediction of soil properties using environmental correlation. Geoderma 89: 67-94.

Moseby, K. E., B. M. Hill, and J. L. Read. 2009. Arid recovery – a comparison of reptile and small mammal populations inside and outside a large rabbit, cat and fox-proof exclosure in arid South Australia. Austral Ecology 34:156-169.

Munday, B. 2017. Those Wild Rabbits: How They Shaped Australia. Wakefield Press Pty, Limited.

National Land & Water Resources Audit and Invasive Animals Cooperative Research Centre. 2008. Assessing Invasive Animals in Australia 2008, NLWRA, Canberra.

Osborne, P. E., and P. J. Leitão. 2009. Effects of species and habitat positional errors on the performance and interpretation of species distribution models. Diversity and Distributions 15:671-681.

Pedler, R. D., R. Brandle, J. L. Read, R. Southgate, P. Bird, and K. E. Moseby. 2016. Rabbit biocontrol and landscape-scale recovery of threatened desert mammals. Conservation Biology 30: 774–782.

Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2007. Updated world map of the Köppen-Geiger climate classification. Hydrology and Earth System Sciences 11:1633–1644.

Pipino, L. L., Y. W. Lee, and R. Y. Wang. 2002. Data quality assessment. Communications of the ACM 45:211–218.

Pitney Bowes Software. 2017. MapInfo Professional.

Phillips, S. J., and M. Dudík. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Plummer, M. 2003. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling.Proceedings of the 3rd International Workshop on Distributed Statistical Computing, 20–22 Mar 2003, Vienna. Vienna, Austria: Technische Universität Wien.

Pocock, M. J. O., J. C. Tweddle, J. Savage, L. D. Robinson, and H. E. Roy. 2017. The diversity and evolution of ecological and environmental citizen science. PlosOne 12:e0172579.

R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Australia.

Ramsey, D. S. L., S. McPhee, D. M. Forsyth, I. G. Stuart, M. Scroggie, M. Lindeman, and J. Matthews. 2014. Warren re-opening rates following coordinated ripping programs on rabbit populations in Victoria, south-eastern Australia. Wildlife Research 41: 46–55.

Reichard, S. H., and P. White. 2001. Horticulture as a Pathway of Invasive Plant Introductions in the United States. BioScience 51:103.

Robertson, M. P., and N. P. Barker. 2006. A technique for evaluating species richness maps generated from collections data. South African Journal of Science 102: 77-84.

Robertson, M. P., G. S. Cumming, and B. F. N. Erasmus. 2010. Getting the most out of atlas data.

Royle, J. A., and J. D. Nichols. 2003. Estimating Abundance from Repeated Presence–Absence Data or Point Counts. Ecology 84(3): 777–790.

Scanlan, J. C., D. M. Berman, and W. E. Grant. 2006. Population dynamics of the European rabbit (*Oryctolagus cuniculus*) in north eastern Australia: simulated responses to control. Ecological Modelling 196: 221-236.

Scroggie, M. P., Forsyth, D. M., S. R. McPhee, J. Matthews, I. G. Stuart, K. A. Stamation, M. Lindeman, and D. S. L. Ramsey. 2018. Invasive prey controlling invasive predators? European rabbit abundance does not determine red fox population dynamics. Journal of Applied Ecology 55: 2621–2631.

Spencer, P. B. S., and J. O. Hampton. 2005. Illegal Translocation and Genetic Structure of Feral Pigs in Western Australia. The Journal of Wildlife Management 69:377–384.

Stodard, E., and I. Parer. 1998. Colonisation of Australia by the rabbit, *Oryctolagus cuniculus* (L.) CSIRO Division of Wildlife and Ecology Project Report No. 6. CSIRO, Australia.

Tablado, Z., E. Revilla, and F. Palomares. 2009. Breeding like rabbits: global patterns of variability and determinants of European wild rabbit reproduction. Ecography 32: 310-320.

Tablado, Z., E. Revilla ,and F. Palomares. 2012. Dying like rabbits: general determinants of spatio-temporal variability in survival. Journal of Animal Ecology 81: 150-161.

Tait, J. 2005. Conserving Australia's terrestrial biodiversity: priorities for a living continent. Report prepared for WWF-Australia, WWF-Australia Sydney.

Thompson, H. V., and C. M. King. 1994. The European rabbit: the history and biology of a successful colonizer. Oxford University Press, Oxford, NY. 245 p.

Twigg, L. E., T. J. Lowe, G. S. Gray, G. R. Martin, A. G. Wheeler, and W. Barker. 1998. Spotlight counts, site fidelity and migration of European rabbits (*Oryctolagus cuniculus*). Wildlife Research 25: 113-122.

Wells, K., B. W. Brook, R. C. Lacy, G. J. Mutze, D. E. Peacock, R. G. Sinclair, N. Schwensow, P. Cassey, R. B. O'Hara, and D. A. Fordham. 2015. Timing and severity of immunizing diseases in rabbits is controlled by seasonal matching of host and pathogen dynamics. Journal of the Royal Society Interface 12: 20141184.

Wells, K., P. Cassey, R. G. Sinclair, G. J. Mutze, D. E. Peacock, R. C. Lacy, B. D. Cooke, R. B. O'Hara, B. W. Brook, and D. A. Fordham. 2016a. Targeting season and age for optimizing control of invasive rabbits. The Journal of Wildlife Management 80: 990-999.

Wells, K., R. B. O'Hara, B. D. Cooke, G. J. Mutze, T. A. A. Prowse, and D. A. Fordham. 2016b. Environmental effects and individual body condition drive seasonal fecundity of rabbits: identifying acute and lagged processes. Oecologia 181: 853-864.

Williams, C.K., I. Parer, B. J. Coman, J. Burley, and M. Braysher. 1995. Managing vertebrate pests: rabbits. Bureau of Resource Sciences/CSIRO Division of Wildlife and Ecology, Australian Government Publishing Service, Canberra.

Yesson, C., P. W. Brewer, T. Sutton, N. Caithness, J. S. Pahwa, M. Burgess, W. A. Gray, R. J. White, A. C. Jones, F. A. Bisby, and A. Culham. 2007. How global is the global biodiversity information facility? PlosOne 11:e1124.