Models of spatiotemporal variation in rabbit abundance reveal management hotspots for an invasive species

Stuart C Brown[1*], Konstans Wells[2], Emilie Roy-Dufresne[1], Susan Campbell[3], Brian Cooke[4], Tarnya Cox[5], Damien A. Fordham[1]

[1] The Environment Institute and School of Biological Sciences, University of Adelaide, Adelaide, South Australia 5005, Australia

[2] Department of Biosciences, Swansea University, Swansea SA2 8PP, UK

[3] Biosecurity and Regulation, Primary Industries and Regional Development, Albany, Western Australia 6330, Australia

[4] Institute for Applied Ecology, University of Canberra, Canberra, ACT 2601, Australia

[5] Vertebrate Pest Research Unit, NSW Department of Primary Industries, Orange, New South Wales 2800, Australia

* Corresponding author. Email: s.brown@adelaide.edu.au

Stuart C Brown: s.brown@adelaide.edu.au; Konstans Wells: k.l.wells@swansea.ac.uk; Emilie Roy-Dufresne: emilie.roy-dufresne@adelaide.edu.au; Susan Campbell: susan.campbell@agric.wa.gov.au; Brian Cooke: brian.cooke@canberra.edu.au; Tarnya Cox: tarnya.cox@dpi.nsw.gov.au; Damien A. Fordham: damien.fordham@adelaide.edu.au

Abstract:

The European rabbit (*Oryctolagus cuniculus*) is a notorious economic and environmental pest species in its invasive range. To better understand the population and range dynamics of this species, 41 years of abundance data have been collected from 116 unique sites across a broad range of climatic and environmental conditions in Australia. We analyzed this time series of abundance data to determine whether inter-annual variation in climatic conditions can be used to map historic, contemporary, and potential future fluctuations in rabbit abundance from regional to continental scales. We constructed a hierarchical Bayesian regression model of relative abundance that corrected for observation error and seasonal biases. The corrected abundances were regressed against environmental and disease variables in order to project high spatiotemporal resolution, continent-wide rabbit abundances. We show that rabbit abundance in Australia is highly variable in space and time, being driven primarily by inter-annual variation in temperature and precipitation in concert with the prevalence of a non-pathogenic virus. Moreover, we show that inter-annual variation in local spatial abundances can be mapped effectively at a continental scale using highly resolved spatiotemporal predictors, allowing "hotspots" of persistently high rabbit abundance to be identified. Importantly, cross-validated model performance was fair to excellent within and across distinct climate zones. Long-term monitoring data for invasive species can be used to map fine-scale spatiotemporal fluctuations in abundance patterns when accurately accounting for inherent sampling biases. Our analysis provides ecologists and pest managers with a clearer understanding of the determinants of rabbit abundance in Australia, offering an important new approach for predicting spatial abundance patterns of invasive species at the near-term temporal scales that are directly relevant to resource management.

Keywords:

climate drivers, invasion hotspot, invasive species management, long-term monitoring, N-mixture model, *Oryctolagus*, random forests

Introduction:

Long-term datasets are vitally important for the successful management of native and invasive species (Yoccoz et al. 2001). In a macro-ecological context, such datasets should ideally report the abundance of the species through time at multiple sites, capturing a broad range of niche requirements (Magurran et al. 2010). These types of spatiotemporal explicit datasets, whilst being rare (Whittaker et al. 2005), can help to identify climatic, environmental and biotic processes that determine species range limits and drive the structure and dynamics of geographic ranges (Caughley et al. 1988, Gaston 2003). Furthermore, they can be used to predict spatiotemporal variation in population fluctuations which is crucial for effective conservation (Channell and Lomolino 2000) and pest and invasive species management (Strayer et al. 2006).

Differences in duration, methodology, frequency of sampling, spatial grain and spatial extent within and between studies, makes systematic comparisons of inferences of long-term population trends, from multiple studies problematic (Magurran et al. 2010, Pagel et al. 2014). Notwithstanding these difficulties, the integration of long-term spatially explicit datasets from disparate sources, with varying spatial and temporal scales, is becoming more common in macroecological studies as a means of forecasting patterns of spatial abundance in time and space (Mellin et al. 2012), and to better understand ecological impacts of invasive species (Strayer et al. 2006). The development of statistical approaches that allow sampling and temporal biases (that result in false variation; Martin et al. 2005), among other sources of uncertainty in abundance time series data (e.g., imperfect detection; Joseph et al. 2009), to be identified and explicitly modelled, is facilitating better estimates of spatiotemporal variation in local-to-regional species abundance patterns (Magurran 2007, Joseph et al. 2009, Schurr et al. 2012).

Although abundance data are still typically analyzed using the average or maximum count per unit transect (Buckland et al. 2015), N-mixture models (models with a mixed binomial-Poisson likelihood), which estimate relative abundance based on replicated counts (Royle 2004), are being used with increasing frequency to explicitly account for detection probability in spatially explicit estimates of abundance (Aubry et al. 2012, Guillera-Arroita 2017), including invasive species (Wells et al. 2016b). Such models essentially produce a bias-corrected abundance estimate, that can then be used to examine trends in population fluctuations and abundances. To do this, corrected abundance data are combined with environmental and climatic information, and where available, data on biotic interactions (i.e. covariate effects; Royle 2004), to determine potential drivers of abundances. These statistical relationships can then be used to make projections of abundance, providing opportunities to assess population trends and geographic range limits across space and time (Joseph et al. 2009, Schurr et al. 2012).

The abundances and subsequent ecological impacts of invasive species are exceedingly variable through time in their invasive range (Sofaer et al. 2018), making them highly suited for N-mixture modelling approaches (Joseph et al. 2009, Guillera-Arroita 2017). For example, the occupancy and abundance rates of invasive species can change over short time scales as a result of shifting evolutionary pressures (Phillips et al. 2006, Cox 2013), modifications in community composition (Mutze et al. 2016), and changes in species interactions (Lurgi et al. 2018). This temporal variability means that invasive species are typically not in equilibrium with their environment (Elith et al. 2010). Therefore, the collection of long-term count data at spatiotemporal resolutions that allow for acute and chronic effects of abiotic and biotic conditions on abundance to be teased apart using robust statistical approaches is vital for generating estimates of population variation for invasive species. These estimates will lead to better ecological and economic outcomes if the mechanisms that affect population fluctuations of invasive species can be identified and managed accordingly (Strayer et al. 2006).

Abundance, not probability of occurrence, is the currency of invasive species management. This is because ecological mechanisms of range limits and shifts can be better

understood using abundance data (Fordham et al. 2012) and spatiotemporal estimates of abundance and subsequent variability are frequently used to maximize the efficacy of on-the-ground management efforts to control invasive species (Simberloff 2003). For example, efforts for managing rabbits in their invasive range are often reduced when population size is perceived to be relatively low, allowing resources to be prioritized elsewhere (Wells et al. 2016a). Moreover, time series abundance estimates are important for the development of spatially explicit and dynamic population models for invasive species management (Mellin et al. 2016), enabling management intervention strategies to be directly tested (Fordham et al. 2013a). This is because time series abundance estimates enable key demographic rates to be estimated, including finite rates of population growth (Hone 1999), and for models to be calibrated (Cabral and Schurr 2010). Therefore, modelling spatiotemporal abundance patterns of invasive species provides resource managers with some distinct advantages over modelling occurrence.

The European rabbit (*Oryctolagus cuniculus*) is an invasive pest species in several regions around the world (Cooke 2012); but is a threatened species in its native range on the Iberian peninsula, where it is a multifunctional keystone species that engineers habitat and provides prey to consumers (Delibes-Mateos et al. 2008, Fordham et al. 2013b). In Australia, *O. cuniculus* occurs across > 70 % of the mainland (Stodart and Parer 1988) and is well adapted to the broad range of climatic conditions of the central and southern parts of the continent (Cooke et al. 2018a), making it one of Australia's most abundant invasive vertebrate species (Cooke 2012). It negatively effects Australia's unique biodiversity directly through outcompeting native species for food and shelter (Cooke 2012), altering vegetation composition (Mutze et al. 2016), and, indirectly, by maintaining invasive predator populations (Lurgi et al. 2018). Furthermore, rabbits adversely affect the agro-economy with an attributed economic loss of approximately A$200m annually (Cooke et al. 2013).

Recently, more than 50 years of historical and contemporary rabbit abundance survey data, from across the invasive range of rabbits in Australia, were combined with high resolution weather, climate and environmental information to create the Australian National Rabbit Database (Roy-Dufresne et al. 2019a). This dataset provides new and important

opportunities to better manage rabbits in Australia through a stronger understanding of their population fluctuations and range dynamics. To date, there has been no Australia-wide assessment of spatiotemporal rabbit abundances and fluctuations. Such an assessment is needed to identify where, and under which environmental conditions rabbits attain high and stable abundances, due to population recruitment and resistance to population perturbations. Using this information in predictive models to generate high spatiotemporal resolution maps of rabbit abundance will enable more targeted rabbit management at local-to-regional scales, enabling the control of invasive rabbits to potentially be better optimized (Fordham et al. 2012, Lurgi et al. 2016). This is because invasive species management decisions are typically based on populations in specific habitats and locales, involving time-horizons of seasons to decades (Rout and Walshe 2013, Wells et al. 2016a).

Here we combine spotlight count data from 116 sites collected over 41 years (1972 – 2012 inclusive) for mainland Australia for *O. cuniculus* (Roy-Dufresne et al. 2019a). Our focal period starts approximately 20 years after the introduction of myxomatosis – a highly infectious and often fatal rabbit disease caused by the myxoma virus – to mainland Australia in 1950. The focal period cover the introduction of Rabbit Hemorrhagic Disease Virus (RHDV) in 1995 (Pech and Hood 1998) and its establishment in rabbit populations. We apply a multi-level hierarchical N-mixture model to these data to "correct" for important biases (namely detection error and bias due to seasonal reproduction and abundance fluctuations) in annual estimates of abundance; and then use machine learning algorithms to determine the relative influence of climate, environment and disease on spatiotemporal variation in rabbit abundance in Australia. Our two-step approach provides a computationally efficient method for generating range-wide, high-resolution spatially (~5 km) and temporally (annual) explicit projections of abundance for a well-established invasive species, which can then be used to reveal hotspots of abundance and to elucidate the drivers of the structure and dynamics of their geographic ranges. Such information is needed to effectively management invasive species at local-to-regional scales through targeted control measures (Baker and Bode 2016, Lurgi et al. 2016).

Methods:

The key steps used to analyze time series count data for rabbits and predict spatiotemporal 'hotspots' of abundance are shown schematically in Figure 1.

Spotlight data:

The spotlight data collated in the Australian National Rabbit Database covers the broad range of climatic and environmental conditions that rabbits inhabit in Australia (Figure 2). These data were collated from several different sources, mostly regional researchers affiliated with state-level governmental institutions. We only consider data records with complete information regarding: (1) geolocation; (2) survey dates; (3) transect length (limited to plausible values ≥ 1 and ≤ 100km); and (4) rabbit encounters (0 counts were included). This resulted in 116 unique grid-cell sites (result of aggregating all count data into 5 km^2 grid cell units) which we subsequently used to model rabbit abundances in space and time.

Climate and environmental data

The aggregated site data was matched to gridded 5 km^2 data containing information on total daily precipitation (mm), and minimum and maximum temperature (°C) data from the TERN e-MAST facility (http://portal.tern.org.au). These climatic variables have been shown to influence the burrow emergence behavior of rabbits, and therefore affect detection probability of rabbits during spotlight surveys (Ballinger and Morgan 2002). Furthermore, grid-cell information on the location and length of the transect, annual climatic conditions (e.g. previous summer total precipitation, previous winter average minimum temperature), environmental conditions (e.g. proportion of agricultural land within the 5km^2 grid cell, proportion of clay soils), and disease proxies (e.g. prevalence of non-pathogenic rabbit calicivirus [RCV-A1]; Liu et al. 2014) was also joined to the aggregated data.

Full details of the climatic and environmental variables used in our analysis are provided in Appendix S1: Table S1.

Climate zone

To account for regional differences in seasonal rabbit population growth and recruitment (Poole 1960, Hone 1999, Mutze et al. 2002, Wells et al. 2016b), the spotlight records were

grouped into climate zones, where long-term climatic conditions were similar. No geophysical properties (e.g. elevation) were included in the stratification units as they are not expected to significantly influence rabbit distributions and abundance in Australia (Cooke 1977, Fordham et al. 2012, Wells et al. 2016b, Roy-Dufresne et al. 2019b). The climate zones were identified by first doing a principal component analysis on 12 climatic variables (e.g. 30-year mean maximum temperature, precipitation seasonality), and then an iterative test to identify the number of "climate zones" required to adequately describe the diversity of climatic conditions. Using a k-means clustering algorithm (R Core Team 2017), the iterative test involved minimizing the within-cluster sum of squared errors (SSE) for a range of potential cluster solutions (2 – 12 clusters) and identifying the point of inflection where the SSE levelled off. Since rabbit reproduction and regional population fluctuations are often highly seasonal (Mutze et al. 2002, Wells et al. 2016b), the spotlight abundance data was grouped by climatic zones where long-term average climate and resource availability (Noble 1977) is assumed to be similar (Figure 2). In doing so, we assume that all rabbit populations from a climate zone follow similar seasonal population fluctuations through synchronous timing of reproduction. Climate zones ranged from warm, humid summer conditions in the north with austral summer (DJF) dominated rainfall (Zone 8) to mild summers and cool winters, with winter dominated rainfall conditions in the south (Zone 5). A full description of the climate zones is provided in Appendix S1: Table S2, with mean (± S.D) for each variable. The climate zones are available in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc.

Correcting abundance count data:

We established a generalized multi-level hierarchical mixture model in a Bayesian framework to derive an index of standardized relative yearly rabbit abundance free of well-established and important sampling biases that is comparable across surveys. Specifically, we used an *N*-mixture model (Royle 2004) to model spotlight counts *c(t,k)* for each of the repeated surveys *k* during a given year in for cell *r* at time step *t* as:

$$c(t,k) \sim Bin[N_r(t), p(t,k)] \; and \; N_r(t) \sim Pois[\Lambda(r,t)] \quad (1)$$

where, we assumed each count to be a random draw based on the underlying true but unknown abundance $N_r(t)$, governed by the Poisson density $\Lambda(r,t)$.

We modelled the detection probability $p(t,k)$ with a *logit*-link function as:

$$logit[p(t,k)] = \mu(r) + \eta(c) + v_1(ds) + v_2(t_{max}) + v_3(t_{min}) + v_4(pr_{day}) \quad (2)$$

where, $\mu(s)$ accounts for variation in detection probability between cells (r), $\eta(c)$ accounts for variation between climate zones (c), $v(ds)$, $v(t_{max})$, $v(t_{min})$, and $v(pr_{day})$ allow for variation between different data sources (corresponding to different survey teams), and daily weather variables. Note that $\Lambda(r,t)$ represents the relative local abundance at the time of observation and is not corrected for transect length nor seasonal bias in counts.

The spotlight surveys were conducted in different seasons and along transects of different lengths, therefore, the estimate of detection probability needs to account for spatiotemporal variability in recruitment (Wells et al. 2016b) and survey efforts (Ballinger and Morgan 2002, Barrio et al. 2009). Doing so will reduce potentially important sources of bias in estimates (Link et al. 2018). We used random regression models, following Wilson et al. (2005), to account for potentially important variation due to seasonal bias and different transect lengths. For this, we modelled $\Lambda(r,t)$ on a log-scale (natural logarithm):

$$\ln[\Lambda(r,t)] = \omega(r,y,m) + \ln[transL(i)] \quad (3)$$

where, $\omega(\mathrm{r}, \mathrm{y}, \mathrm{m})$ is the intercept of the random regression model, which can vary across cells (r), years (y), and months (m); and $transL(i)$ is the corresponding length of the transect in km for the respective sample.

We modelled $\omega(\mathrm{r}, \mathrm{y}, \mathrm{m})$ further based on a two-level mean/hyperprior models with Gaussian error of the general form $\mathrm{N}[H, \sigma^2]$ with mean H and variance σ^2 in order to derive an estimate of the most likely yearly abundance independent of the seasonal bias in counts:

$$\omega(r,y,m) \sim \mathrm{N}[H_{RY.\omega}(r,y), \sigma_{R.\omega}(r)^2] \quad (4)$$

and

$$H_{RY.\omega}(r,y) \sim \mathrm{N}[H_{R.\omega}(r), \sigma_{R.\omega}(r)^2] \quad (5)$$

where, $\mathrm{H_{RY.\omega}(r,y)}$ denotes the population-level annual 'average' density distribution for each site and year, while the respective variance $\sigma_{\mathrm{R.\omega}}(\mathrm{r})^2$ is a random estimate of the inter-seasonal variance in population fluctuations for each cell (r) according to the variation in

monthly counts from different sites. $H_{R.\omega}(r)$ denotes the site-level 'average' relative abundance and $\sigma_{R.\omega}(r)^2$ interannual variance in population fluctuations.

Our model was fitted in a Bayesian framework using Markov Chain Monte Carlo (MCMC) sampling in the software JAGS 4.3.0, operated via the R software (v 3.5.1; R Core Team 2017) with the 'rjags' package (Plummer 2016). Priors were specified as $\sigma \sim$ Gamma (0.01, 0.01) and $H_{R.\omega}(r) \sim N(0, 10)$.

We applied posterior predictive model diagnostics to assess whether the model assumptions were good approximations of the data generating process. Bayesian *P*-values ~0.5 indicate a good model fit whereas values close to 0 or 1 indicate an increasing discrepancy between model predictions and observation data (Gelman et al. 1996). We calculated 5000 values as posterior distributions for all parameters of interest after mixing and convergence of two parallel MCMC chains was confirmed visually and with the Gelman-Rubin diagnostic (most values < 1.2) after a burn-in of 50,000 MCMC samples.

Modelling rabbit abundances in space and time:

Modelled corrected abundances ($H_{RY.\omega}(r, y)$; Appendix S1: Figure S1) – extracted from the Bayesian model as the mode of the posterior density distributions – were aligned with the key predictor variables in Appendix S1: Table S1. To generate predictions of rabbit density we opted to use a readily available and straight forward machine learning approach, random forests (RF; Breiman 2001), implemented in the 'randomForest' package for R (Liaw and Wiener 2018). This decision was made because compared to most Bayesian methods, RFs enable a broad range of model selection and cross-validation procedures that are not commonly used in an MCMC context due to practical computing time issues (Kéry and Schaub 2011). Essentially, we use the hierarchical model for correcting abundances for various sources of bias, while we use RFs for exploring and accounting for any non-linear patterns between corrected abundances and climate and environmental correlates in spatial predictions.

For simplicity, we categorized the annual model corrected abundance estimates into 'low density' (corresponding to ≤ 45 rabbits/km^2, assuming an effective transect width of 0.1

km) and 'high density' (corresponding to > 45 rabbits/km^2). These values were chosen as there was a natural break in the calculated density values, and it has been demonstrated that at densities < 50 rabbits/km^2, damage to perennial native vegetation is minimized across both temperate and arid environments (Mutze et al. 2016). This binary variable allowed us to use a two-step regression model (Mellin et al. 2012) to predict rabbit abundance through time and space (Figure 1). To do this we used a binary classification variable (high/low density) as a predictor in our regression model in order to make spatiotemporal explicit predictions of rabbit abundance. Thus, the model generates spatial estimates of abundance conditional on whether density in a given cell is high or low.

A binary classification RF model was first created using the aforementioned density classes. Due to the unbalanced nature of the classes *a posteriori* tuning of the probability threshold for classification (default = 0.50) was performed to maximize model accuracy. Using this approach, once the classification RF model had been trained (using the default threshold), a series of thresholds between 0 and 1 were tested to maximize the area under the ROC curve (Kuhn 2018). A separate regression RF model was then built which incorporated the RF predicted density classes as an additional correlate for predicting rabbit abundance. The two-step procedure necessitated the use of RF predicted rabbit density for cell *r* at time t_{-1}.

To train and assess the performance of our RF models, we used 1,000 repeats of a Monte-Carlo cross-validation procedure, whereby the data was randomly split (without replacement) into 80/20 training/test splits. Models consisting of 2001 trees were built on the 80% splits and tested on the 20% holdout test samples. The cross-validation was carried out using the 'caret' package for R (Kuhn 2018). Correlations between the environmental and climatic correlates (Appendix S1: Table S1) were tested, and with a few exceptions, were not strongly positive or negatively correlated (Appendix S1: Figure S2). As random forests are able to deal with correlated variables (Breiman 2001), we did not perform any feature selection or reduction on the correlates included in the model. Variable importance was calculated using an unscaled permutation-based method (Liaw and Wiener 2018). Spatial autocorrelation in the residuals of predicted RF abundance was conducted using Moran's I

statistic using an inverse distance relationship with a threshold of 1130 km (the average pairwise distance between all sampling points).

Spatially and temporally explicit estimates of rabbit density and abundance (i.e. maps), were then predicted for Australia between 1972 and 2012. Hotspots of high rabbit abundance were identified for each year that abundance predictions were made by approximating a cumulative relative frequency distribution (CRFD) curve as a function of relative rabbit abundance plotted against the relative frequency distribution of abundance for a given year (Bartolino et al. 2011a). The count data for the CRFD curve was pooled across all sites within each year. Unlike traditional hotspot identification methods which require a user-defined threshold (Bartolino et al. 2011a, b, Cayuela et al. 2011), the CRFD method identifies the hotspot threshold as a function of the CRFD curve. The method involves calculating the slope of the tangents to the CRFD curve, with the hotspot threshold being identified as the value where the slope of the tangent is 45° (Bartolino et al. 2011a). Using the thresholds for each year, the predicted abundance maps were then converted to a series of binary hotspot maps. The proportion of time spent in "hotspot" conditions across the study period was then determined for each cell by counting the number of times a cell was classified as a hotspot, divided by the number of years.

Climate envelope and confidence mask

To constrain predictions of corrected density and abundance to climate conditions used to train the model and, in doing so, avoid model extrapolation, we identified all cells for Australia that fell within the extremes (i.e. highest and lowest) of the climatic variables that were used to define the climate zones. In doing so we constrain our predictions of density and abundance to cells that fall within the bounds of our long-term climatic variables and limit extrapolation to "unseen" climatic conditions (Liu et al. 2014). All maps in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc are clipped to this climate envelope.

To identify regions of "low confidence," we did a PCA using the top 10 environmental and climatic variables that were considered important for both the classification and regression RF models, and using the first 3-components as pseudo-3-

dimensional-coordinates, we identified individual grid cells in 3-dimensional space that were outside a 3-dimensional convex-hull of the sampling sites within each of the climate zones. That is, 3-dimensional polygons were constructed, with vertices defined by the pseudo-3-dimensional-coordinates of all cells for which we had samples within each climate zone. Any additional cells (i.e. grid cells where no monitoring had occurred) that exist inside the 3-dimensional boundary of this polygon can be considered inside the "environmental space" of the study sites within each climate zone. Cells outside this 3-dimensional polygon, have climatic or environmental conditions not seen at the sites used within the climate zone. This process allowed us to identify grid-cells where our predictions of corrected density and abundance were outside the environmental and climatic space of all monitoring sites in the study, essentially creating a mask of low confidence where predictions were likely to have been extrapolated in environmental and/or climatic space, rather than interpolated. The confidence mask is supplied in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc.

Results:

Drivers of detection

Coefficient estimates for the drivers of detection varied in magnitude and sign across climate zones (Table 1). Within 5 of the 9 climate zones (1, 2, 4, 6, and 7) an increase in maximum daily temperature decreased the probability of detection (mean = -0.23, S.D. = 0.11), whilst for the remaining 4 zones (3, 5, 8, and 9) it resulted in an increase in detection probability (mean = 0.28, S.D. = 0.18). Increases in minimum daily temperature (except for climate zone 4) and daily total precipitation, resulted in slight increases in detection probability (minimum daily temperature: mean = 0.04, S.D. = 0.03; precipitation: mean = 0.05, S.D. = 0.02).

Gelman-Rubin diagnostics for hyperpriors indicated good convergence of the detectability model: (i) maximum daily temperature = 1.08 (upper CI = 1.32); (ii) minimum daily temperature = 1.17 (upper CI = 1.59); (iii) daily precipitation = 1.09 (upper CI = 1.31).

Determinants of spatiotemporal patterns of density and abundance

The five most important predictor variables, as calculated using permutation importance, for the classification RF model of rabbit density (high/low) were: temperature seasonality,

proportion of irrigated agriculture within the grid-cell, two-year lagged average temperature, distance to permanent water, and the prevalence of the benign RCV-A1 calicivirus (Appendix S1: Figure S3). Evaluation of these variables with Accumulated Local Effects (ALE; Apley 2016) plots (Appendix S1: Figure S4) suggests that at sites with lower temperature seasonality and lower mean annual temperatures in the years (n = 2) prior to the survey, there was an increased probability of high rabbit densities. Furthermore, at the few sites that were sampled on irrigated land, sites with as little as ~7.5% irrigated agriculture within a 5km grid-cell were 60% more likely to have high rabbit densities than a conditional averaged site (Appendix S1: Figure S4). The probability of a site having high density decreased with distance from permanent water until ≳ 20km. Although some sites that had high densities were situated very far away from permanent water sources (> 150km; Appendix S1: Figure S4). As benign RCV-A1 prevalence increased above ~20% the probability of a site having high abundance was ~25% greater than the average site (Appendix S1: Figure S4). Taken together the results from the classification RF model suggest that at sites with stable, cool temperatures, with abundant pastures (owing to irrigation, and proximity to permanent water sources), and prevalent RCV-A1, rabbit densities will be higher. Cross-validated performance of the classification model was excellent across climate zones and in most cases fair to excellent within climate zones (Table 2). The mode of the predicted rabbit densities and the mean prediction probability confidence (1972 – 2012) is shown in Figure 3. The complete time series of density maps is shown in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc.

The five most important variables influencing model predictions of rabbit abundance (excluding rabbit density class from the previous year; Appendix S1: Figure S3) were: precipitation the year prior, average winter minimum temperature the year prior, two-year lagged total precipitation, precipitation seasonality, and the prevalence of benign RCV-A1. These variables exhibited similar importance within the regression RF model (Appendix S1: Figure S3). ALE plots of the regression model (Figure 4) suggest that at sites where one-year lagged precipitation was ≳ 400mm the year prior, and one-year lagged minimum winter temperatures ≳ 4°, there was an increase in rabbit abundance. As for the classification model, examination of two year lagged average temperature suggests that at sites with lower mean

annual temperatures (≲ 16° C) rabbit abundances are higher. Like with the classification RF, this suggests that in cool, wet areas rabbit abundances will be higher. However, for the regression RF the effects of precipitation are likely acting as a proxy for pasture availability. In contrast to the classification model, the plot of the association between abundance and the prevalence of benign RCV-A1 suggests that as virus prevalence increases, abundances will decrease (Figure 4). Regression model performance was excellent across climate zones and in most cases fair to excellent within climate zones (Table 1). With some exceptions, the 95% prediction intervals of the regression RF model were able to capture the variability in the corrected abundance counts (Figure 5). Moran's I suggested no residual spatial autocorrelation ($I = 0.002$, $z = 0.166$, $p = 0.87$). The mean of the predicted rabbit abundances and the mean standard deviation of the abundances through time (1973 – 2012) are shown in Figure 3. All abundance maps are available in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc.

Variable interactions

Examination of second-order interactions for the regression model (Figure 4b) suggest that for sites with > 20% prevalence of benign RCV-A1 and with mean annual temperatures below ~16° C, the combined effects are positive on abundance. Conversely, second-order interactions between precipitation and RCV-A1 (Figure 4c) suggest that at sites with ≲ 20% prevalence, and annual precipitation is > 600mm/year, there is a net positive effect on rabbit abundance suggesting increased precipitation, and therefore an increase in potential pasture availability, could lead to higher abundances at sites where there is low immunity to lethal RHDV. Furthermore, after the main effects of the variables (namely precipitation and RCV-A1 prevalence) have been accounted for, there is a negligible effect of RCV-A1 prevalence on rabbit abundance in areas of high rainfall (> 600mm/year) as illustrated by the relative "flatness" and values of the interaction surface (Figure 4c).

Hotspots of abundance

Predicted maps of rabbit density and abundance show high levels of interannual variation along the east coast of Australia, with density classes and abundances fluctuating between

high and low levels inland from the coast for a distance of ~300km (Figure 3). Hindcasts of abundance for the southern coastline of mainland Australia showed lower inter-annual variability compared to the eastern coastline. Inland areas tended to be relatively stable, however, these areas were generally inside the low-confidence mask.

Analysis of temporal hotspot conditions suggest small inter-annual variability in the threshold used to identify hotspots of rabbit abundance across time (mean = 12, S.D. = 2.5; Figure 6). A Mann-Kendall trend test suggested a very small but non-significant increase in the hotspot threshold across all years ($\tau = 0.05$, $p = 0.70$) suggesting rabbit numbers are decreasing as the threshold used to identify a hotspot has increased. Broadly, hotspots of rabbit abundance followed similar patterns to maps of predicted abundance (Figure 6). All hotspot maps are provided in Data S1: Brown_et_al_RabbitAnalysisOutputs.nc, with code to reproduce the analysis and Figure 6 provided in Data S1: RabbitHotpotAnalysis.R.

Discussion:

We show that combining multi-level hierarchical Bayesian models with simple machine learning methods is an effective, and computationally tractable, method for modelling the structure and dynamics of the geographical distribution of a wide-ranging invasive vertebrate species. By allowing us to project, effectively, spatiotemporal variation in the abundance of rabbit populations across their invasive range in Australia, our approach provides an important new approach for identifying hotspots of persistently high rabbit abundance at timescales (annual to decadal) that are directly relevant to resource management in Australia (Hobday et al. 2010, Cooke 2012). The identification of potential past and future hotspots will allow pest managers to better plan for and conduct control actions for rabbits in areas and at times of greatest need (Simberloff 2003).

A robust understanding of spatiotemporal variability in abundance patterns across an invasive species range is invaluable for studying the processes of invasion and for pest management (Strayer et al. 2006). However, this requires spatial abundance estimates that account for inherent biases in long-term count data (Joseph et al. 2009, Link et al. 2018). Our use of a hierarchical Bayesian N-mixture model, coupled with a hyperprior-based model,

allowed us to explicitly account for important variation in detection probability in long-term estimates of abundance from sites covering a large spatiotemporal extent. Although similar approaches have been applied to rabbit spotlight count data before (Aubry et al. 2012, Wells et al. 2016b), they have not then been extended to account for seasonal bias in counts from multiple surveys at multiple sites, and have not been used to for predictions of spatiotemporal abundance. Daily precipitation has previously been found to drastically decrease surface activity of rabbits and therefore detectability (Ballinger and Morgan 2002), whilst the results of our analysis suggest very small increases in detectability with increased rainfall on the day of the survey. We found that increases in daily maximum temperature could either increase or decrease the probability of detection across climate zones – this variability could be attributed to differences in climate, or differences in survey time. For example, in climate zone 9 (warm, semi-arid conditions; Appendix S1: Table S2) increases in maximum daily temperature increased detection probability, but if these surveys were mainly conducted in the Austral winter (June, July, August) then maximum daily temperatures would be lower than in, for example, climate zone 8 (warm, summer rainfall; Appendix S1: Table S2) (where increases in maximum temperature decreased detection probability) if surveys in that zone were mainly conducted in the Austral summer (DJF). Time of day and year has been shown to influence counts during spotlight surveys (Ballinger and Morgan 2002, Williams et al. 2007, Barrio et al. 2009), with seasonal variations in Mediterranean environments also reported (Martins et al. 2003). These variations have largely been attributed to changes in foraging activities and seasonal variations in population abundances. Whilst information on spotlight survey time (dawn, dusk, hour etc.) is missing from the database, we attempted to overcome seasonal variability by using random regression models (Wilson et al. 2005) which accounted for the month of survey mainly to control for seasonal birth pulsing (Wells et al. 2016b).

Our Bayesian hierarchical N-mixture model, employed at multiple sites across the invasive range of rabbits in Australia, generated a standardized relative yearly rabbit abundance across space and time, corrected for 'false variation' as a result of imperfect detection and seasonal fluctuations in abundance (due to seasonal birth pulses) in some areas.

Using these corrected abundances in a machine learning framework (random forest models; RF) provided a highly computationally efficient method for identifying the spatiotemporal drivers of heterogeneity in rabbit abundance across their invasive range. Furthermore, the use of RF permitted the analysis of non-linear relationships between, and within, predictor variables, that may have otherwise been missed (Breiman 2001, Apley 2016).

The comprehensive nature of the statistical modelling undertaken here, covering 41 years and approximately 3.8 million square kilometers, which is unique for a terrestrial invasive vertebrate, allowed us to directly model potential drivers of rabbit abundance in space and time. Our results support previous research, showing that rabbit populations are influenced by pasture availability and the water content of pastures (Cooke 2012), temperature (attaining higher densities in cooler temperatures; Gilbert et al. 1987) and to an extent, the effects of RCV-A1 (Cooke et al. 2018b). The effects of pasture availability and water content (e.g. the presence of irrigated pastures), combined with cooler temperatures and stable seasonal rainfall could potentially be driving the high predicted densities inland along the southern coastline (Figure 3). The classification RF shows that RCV-A1 prevalence was positively correlated with high rabbit densities (high/low), because RCV-A1 provides partial protection from lethal RHDV (Strive et al. 2013). Since the prevalence of RCV-A1 is strongly correlated with rainfall (Liu et al. 2014), rabbits in cooler, wetter areas are likely to be found in higher abundances due to food availability, and these same populations are likely to be more resistant to infection with lethal RHDV. Although, RCV-A1 also influenced estimates of abundance from the regression RF, prevalence of RCV-A1 was negatively associated with rabbit abundance – a result likely due to complex interactions between rabbit populations, climatic conditions, disease vectors, and epizootics (Cooke 1983, Liu et al. 2014). This result requires further investigation, using new approaches that can directly tease apart these interactions, which are often mutually reinforcing (Wells et al. 2018). The presence of European rabbit fleas was not important in determining either density or abundance, but has been shown previously to decrease rabbit numbers as a vector of the myxoma virus (Cooke 1983). These results are likely a result of a lack of spatial and

489 temporally explicit data on the release, establishment and expansion of European fleas over
490 the study period.
491 Spatial and temporally explicit estimations of abundance are an important component
492 in ecological modelling, because a change in population abundance over time and space is
493 brought about by changes in movement and/or birth and death rates. Identifying and
494 quantifying the temporal variability in populations of invasive species can help to better
495 target management actions (Simberloff 2003) – particularly when responses to removal of
496 invasive species happen on varying time scales (Strayer et al. 2006, Cooke 2012, Lurgi et al.
497 2018), and different control measures will have differing population level effects (Wells et al.
498 2016a). Just as importantly, quantifying variability in population abundances across the range
499 of a species can unlock the ecological mechanisms responsible for spatiotemporal variation in
500 abundance across a species geographic distribution (Caughley et al. 1988). This information
501 is essential for optimizing management actions aimed at controlling invasive species (Mellin
502 et al. 2016). The clustered nature of the spotlight survey data (Figure 2; Roy-Dufresne et al.
503 2019a) resulted in large areas of our predicted density and abundance values falling inside
504 our low confidence mask (Figure 3). However, we still have faith in the predicted values at
505 these sites due to the results of our random forest models, and the climatic mask that was
506 used to prevent any extrapolation in our predictions. Due to the highly fluctuating nature of
507 rabbit populations, we believe that managers should still utilize such information to guide
508 potential future monitoring and management activities. The results also demonstrate that
509 despite using the most extensive dataset available, accurate spatio-temporal predictions of
510 population density and abundance of highly fluctuating species is an ongoing challenge.
511 The concept of a hotspot is focused on the stability of a species population, or a
512 measure of biodiversity, across space and over time (Mittermeier et al. 2011). Since
513 identifying priority areas for control activities is a key component in managing invasive
514 species (Lurgi et al. 2016), our hindcasts of areas where a higher than expected number of
515 rabbits can be found historically (hotspots), reveal priority sites for rabbit management.
516 Efforts towards eradication and control of rabbits in areas that are consistently identified as
517 hotspots (i.e. high proportion of time as a hotspot) should be given top priority (Lurgi et al.

518 2016), and rabbits of all age classes should be targeted (Wells et al. 2016a). Near-term
519 climate forecasts from regional climate models are now readily available at the temporal and
520 spatial resolutions needed to effectively manage biodiversity in response to global change
521 (Tabor and Williams 2010). Integrating these forecasts into our hotspot analysis would allow
522 areas for future rabbit management to be prioritized, based on knowledge of what the
523 abundance pattern is projected to be given forecast climate conditions, with the caveat that
524 inter-warren population competition and structure can occur at finer spatial resolutions than
525 the $5km^2$ spatial scale used in this study (Lurgi et al. 2016). A comparable approach is
526 commonly used in fisheries management to set catch rates in space and time, with good
527 success (Hobday et al. 2010). Furthermore, these same spatiotemporal estimates of density
528 and abundance can be integrated into spatially explicit models for informing pest
529 management (Mellin et al. 2016), examining multispecies interactions (Lurgi et al. 2018), or
530 for inferring patterns of spatial variation in demographic rates (Hone 1999).
531 Beyond rabbits, our method of detecting hotspots could be used to strengthen theory
532 underpinning important concepts in invasion biology, including whether species' populations
533 with high but variable abundance are more sensitive to control interventions than populations
534 with high and less variable abundances (Wells et al. 2016a). The hypothesis being, that
535 populations that are consistent hotspots of abundance, are likely to be more resilient to
536 environmental fluctuations and, therefore, better able to recover from management
537 interventions if not completely eradicated (Mutze et al. 2010). Likewise, the method could be
538 used to test whether hotspots of abundance tend to occur more frequently at the center of the
539 range of an invasive species, as would be expected, based on theory (i.e., species abundance
540 distribution; Lawton 1993)
541 The spotlight count data in the Australian National Rabbit Database includes some
542 studies set up to record trends in rabbit population abundances after active (e.g. warren
543 ripping) or passive (e.g., release of RHDV) population control. The timing and extent of these
544 management actions are not extensively documented, potentially biasing abundance records
545 used in this analysis. While we tried to account for this variation by using data source as a
546 detectability covariate, the absence of detailed information on type and timing of

management intervention prevented a more thorough analysis of this potential source of uncertainty. This bias could possibly have resulted in lower frequencies of high abundance records for some areas, and therefore it is possible that some sites were not correctly identified as hotspots in years directly following management intervention. However, our CRFD hotspot thresholds are well below those shown to influence vegetation community composition (Mutze et al. 2016), and as such any identified hotspots are still valuable for management as these areas would still be identified as priority control sites at abundances lower than those known to alter, for example, vegetation communities.

Our modelling approach for rabbits is applicable to any species for which long-term spatially explicit monitoring data and high temporal resolution climate data are available. It is flexible with regards to spatial grain and extent, and benefits from the utilization of computationally efficient machine learning methods for generating maps of spatial abundance, and for estimating variable interactions that link patterns of abundances to potential drivers and processes. The long-term management of invasive species relies on informed future projections of abundance and a good understanding of the processes that underpin the structure and dynamics of a species past and present-day distribution in its invasive range. Model-based approaches such as those created in this study promise to help guide adaptive management approaches by permitting inferences to be made regarding likely future population trends at scales relevant to management.

Acknowledgements:

This research was funded by an Australian Research Council (ARC) Linkage Project (LP12020024) and Future Fellowship (FT140101192). We would like to thank Greg Mutze for discussions and advice regarding preliminary analysis and outputs.

Literature cited:

Apley, D. W. 2016. Visualizing the effects of predictor variables in black box supervised learning models. arXiv preprint arXiv:1612.08468.

Aubry, P., D. Pontier, J. Aubineau, F. Berger, Y. Léonard, B. Mauvy, and S. Marchandeau. 2012. Monitoring population size of mammals using a spotlight-count-based

575 abundance index: How to relate the number of counts to the precision? Ecological
576 indicators **18**:599-607, doi:10.1016/j.ecolind.2012.01.019.
577 Baker, C. M., and M. Bode. 2016. Placing invasive species management in a spatiotemporal
578 context. Ecological Applications **26**:712-725, doi:10.1890/15-0095.
579 Ballinger, A., and D. G. Morgan. 2002. Validating two methods for monitoring population
580 size of the European rabbit (*Oryctolagus cuniculus*). Wildlife Research **29**:431-437,
581 doi:10.1071/wr01055.
582 Barrio, I. C., P. Acevedo, and F. S. Tortosa. 2009. Assessment of methods for estimating wild
583 rabbit population abundance in agricultural landscapes. European Journal of Wildlife
584 Research **56**:335-340, doi:10.1007/s10344-009-0327-7.
585 Bartolino, V., L. Maiorano, and F. Colloca. 2011a. A frequency distribution approach to
586 hotspot identification. Population Ecology **53**:351-359, doi:10.1007/s10144-010-
587 0229-2.
588 Bartolino, V., L. Maiorano, and F. Colloca. 2011b. Frequency distribution curves and the
589 identification of hotspots: response to comments. Population Ecology **53**:603-604,
590 doi:10.1007/s10144-011-0273-6.
591 Breiman, L. 2001. Random forests. Machine learning **45**:5-32,
592 doi:10.1023/a:1010933404324.
593 Buckland, S. T., E. A. Rexstad, T. A. Marques, and C. S. Oedekoven. 2015. Distance
594 sampling: methods and applications. Springer, Heidelberg.
595 Cabral, J. S., and F. M. Schurr. 2010. Estimating demographic models for the range dynamics
596 of plant species. Global Ecology and Biogeography **19**:85-97, doi:10.1111/j.1466-
597 8238.2009.00492.x.
598 Caughley, G., D. Grice, R. Barker, and B. Brown. 1988. The edge of the range. Journal of
599 Animal Ecology **57**:771-785, doi:10.2307/5092.
600 Cayuela, L., L. Gálvez-Bravo, L. M. Carrascal, and F. S. de Albuquerque. 2011. Comments
601 on Bartolino et al. (2011): limits of cumulative relative frequency distribution curves
602 for hotspot identification. Population Ecology **53**:597, doi:10.1007/s10144-011-0272-
603 7.

604 Channell, R., and M. V. Lomolino. 2000. Dynamic biogeography and conservation of
605 endangered species. Nature **403**:84-86, doi:10.1038/47487.
606 Cooke, B. 1983. Changes in the Age-Structure and Size of Populations of Wild Rabbits in
607 South Australia, Following the Introduction of European Rabbit Fleas, *Spilopsyllus*
608 *cuniculi* (Dale), as Vectors of Myxomatosis. Wildlife Research **10**:105-120,
609 doi:10.1071/WR9830105.
610 Cooke, B. 2012. Rabbits: manageable environmental pests or participants in new Australian
611 ecosystems? Wildlife Research **39**:279-289, doi:10.1071/WR11166.
612 Cooke, B., M. Brennan, and P. Elsworth. 2018a. Ability of wild rabbit, *Oryctolagus*
613 *cuniculus*, to lactate successfully in hot environments explains continued spread in
614 Australia. Wildlife Research **45**:267-273, doi:10.1071/wr17177.
615 Cooke, B., P. Chudleigh, S. Simpson, and G. Saunders. 2013. The Economic Benefits of the
616 Biological Control of Rabbits in Australia, 1950–2011. Australian Economic History
617 Review **53**:91-107, doi:10.1111/aehr.12000.
618 Cooke, B., R. P. Duncan, I. McDonald, J. Liu, L. Capucci, G. J. Mutze, and T. Strive. 2018b.
619 Prior exposure to non-pathogenic calicivirus RCV-A1 reduces both infection rate and
620 mortality from rabbit haemorrhagic disease in a population of wild rabbits in
621 Australia. Transboundary and Emerging Diseases **65**:e470-e477,
622 doi:10.1111/tbed.12786.
623 Cooke, B. D. 1977. Factors limiting the distribution of the wild rabbit in Australia.
624 Proceedings of the Ecological Society of Australia **10**:113-120.
625 Cox, G. W. 2013. Alien Species and Evolution: The Evolutionary Ecology of Exotic Plants,
626 Animals, Microbes, and Interacting Native Species. Island Press, Washington.
627 Delibes-Mateos, M., M. Delibes, P. Ferreras, and R. Villafuerte. 2008. Key role of European
628 rabbits in the conservation of the Western Mediterranean basin hotspot. Conservation
629 Biology **22**:1106-1117, doi:10.1111/j.1523-1739.2008.00993.x.
630 Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species.
631 Methods in Ecology and Evolution **1**:330-342, doi:10.1111/j.2041-
632 210X.2010.00036.x.

633 Fordham, D., H. Akçakaya, M. Araújo, B. Brook, J. Brodie, E. Post, and D. Doak. 2012.
634 Modelling range shifts for invasive vertebrates in response to climate change.*in*
635 Brodie J, Post E, and D. D, editors. Wildlife Conservation in a Changing Climate.
636 University of Chicago Press, Chicago.
637 Fordham, D., H. R. Akcakaya, M. B. Araujo, D. A. Keith, and B. W. Brook. 2013a. Tools for
638 integrating range change, extinction risk and climate change information into
639 conservation management. Ecography **36**:956-964, doi:10.1111/j.1600-
640 0587.2013.00147.x.
641 Fordham, D., H. R. Akçakaya, B. W. Brook, A. Rodríguez, P. C. Alves, E. Civantos, M.
642 Triviño, M. J. Watts, and M. B. Araújo. 2013b. Adapted conservation measures are
643 required to save the Iberian lynx in a changing climate. Nature Climate Change **3**:899,
644 doi:10.1038/nclimate1954.
645 Gaston, K. J. 2003. The structure and dynamics of geographic ranges. Oxford University
646 Press, New York.
647 Gelman, A., X. L. Meng, and H. Stern. 1996. Posterior predictive assessment of model fitness
648 via realized discrepancies. Statistica sinica **6**:733-760.
649 Gilbert, N., K. Myers, B. D. Cooke, J. D. Dunsmore, P. J. Fullagar, J. A. Gibb, D. R. King, I.
650 Parer, S. H. Wheeler, and D. H. Wood. 1987. Comparative Dynamics of Australasian
651 Rabbit-Populations. Wildlife Research **14**:491-503, doi:10.1071/wr9870491.
652 Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and
653 communities under imperfect detection: advances, challenges and opportunities.
654 Ecography **40**:281-295, doi:10.1111/ecog.02445.
655 Hobday, A. J., J. R. Hartog, T. Timmiss, and J. Fielding. 2010. Dynamic spatial zoning to
656 manage southern bluefin tuna (*Thunnus maccoyii*) capture in a multi-species longline
657 fishery. Fisheries Oceanography **19**:243-253, doi:10.1111/j.1365-2419.2010.00540.x.
658 Hone, J. 1999. On rate of increase (r): patterns of variation in Australian mammals and the
659 implications for wildlife management. Journal of Applied Ecology **36**:709-718,
660 doi:10.1046/j.1365-2664.1999.00439.x.

661 Hosmer, D. W., S. Lemeshow, and R. X. Sturdivant. 2013. Applied Logistic Regression. 3rd
662 edition. John Wiley & Sons, Hoboken, NJ, USA.
663 Joseph, L. N., C. Elkin, T. G. Martin, and H. P. Possingham. 2009. Modeling abundance
664 using N-mixture models: the importance of considering ecological mechanisms.
665 Ecological Applications **19**:631-642, doi:10.1890/07-2107.1.
666 Kéry, M., and M. Schaub. 2011. Bayesian Population Analysis using WinBUGS: A
667 Hierarchical Perspective. Academic Press, Oxford.
668 Kuhn, M. 2018. caret: Classification and Regression Training.
669 Lawton, J. H. 1993. Range, population abundance and conservation. Trends in Ecology &
670 Evolution **8**:409-413, doi:10.1016/0169-5347(93)90043-O.
671 Liaw, A., and M. Wiener. 2018. Classification and Regression by randomForest.
672 Link, W. A., M. R. Schofield, R. J. Barker, and J. R. Sauer. 2018. On the robustness of N-
673 mixture models. Ecology **99**:1547-1551, doi:10.1002/ecy.2362.
674 Liu, J., D. A. Fordham, B. D. Cooke, T. Cox, G. Mutze, and T. Strive. 2014. Distribution and
675 prevalence of the Australian non-pathogenic rabbit calicivirus is correlated with
676 rainfall and temperature. PLOS ONE **9**:e113976, doi:10.1371/journal.pone.0113976.
677 Lurgi, M., E. G. Ritchie, and D. A. Fordham. 2018. Eradicating abundant invasive prey could
678 cause unexpected and varied biodiversity outcomes: The importance of multispecies
679 interactions. Journal of Applied Ecology **55**:2396-2407, doi:10.1111/1365-
680 2664.13188.
681 Lurgi, M., K. Wells, M. Kennedy, S. Campbell, and D. A. Fordham. 2016. A Landscape
682 Approach to Invasive Species Management. PLOS ONE **11**:e0160417,
683 doi:10.1371/journal.pone.0160417.
684 Magurran, A. E. 2007. Species abundance distributions over time. Ecology letters **10**:347-
685 354, doi:10.1111/j.1461-0248.2007.01024.x.
686 Magurran, A. E., S. R. Baillie, S. T. Buckland, J. M. Dick, D. A. Elston, E. M. Scott, R. I.
687 Smith, P. J. Somerfield, and A. D. Watt. 2010. Long-term datasets in biodiversity
688 research and monitoring: assessing change in ecological communities through time.
689 Trends in Ecology & Evolution **25**:574-582, doi:10.1016/j.tree.2010.06.016.

690 Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J.
691 Tyre, and H. P. Possingham. 2005. Zero tolerance ecology: improving ecological
692 inference by modelling the source of zero observations. Ecology letters **8**:1235-1246,
693 doi:10.1111/j.1461-0248.2005.00826.x.
694 Martins, H., H. Barbosa, M. Hodgson, R. Borralho, and F. Rego. 2003. Effect of vegetation
695 type and environmental factors on European wild rabbit (*Oryctolagus cuniculus*)
696 counts in a southern Portuguese montado. Acta Theriologica **48**:385-398,
697 doi:10.1007/bf03194177.
698 Mellin, C., M. Lurgi, S. Matthews, M. A. MacNeil, M. J. Caley, N. Bax, R. Przeslawski, and
699 D. A. Fordham. 2016. Forecasting marine invasions under climate change: Biotic
700 interactions and demographic processes matter. Biological Conservation **204**:459-467,
701 doi:10.1016/j.biocon.2016.11.008.
702 Mellin, C., B. D. Russell, S. D. Connell, B. W. Brook, and D. A. Fordham. 2012. Geographic
703 range determinants of two commercially important marine molluscs. Diversity and
704 Distributions **18**:133-146, doi:10.1111/j.1472-4642.2011.00822.x.
705 Mittermeier, R. A., W. R. Turner, F. W. Larsen, T. M. Brooks, and C. Gascon. 2011. Global
706 Biodiversity Conservation: The Critical Role of Hotspots. Pages 3-22 *in* F. E. Zachos
707 and J. C. Habel, editors. Biodiversity Hotspots: Distribution and Protection of
708 Conservation Priority Areas. Springer Berlin.
709 Mutze, G., P. Bird, J. Kovaliski, D. Peacock, S. Jennings, and B. Cooke. 2002. Emerging
710 epidemiological patterns in rabbit haemorrhagic disease, its interaction with
711 myxomatosis, and their effects on rabbit populations in South Australia. Wildlife
712 Research **29**:577-590, doi:10.1071/wr00100.
713 Mutze, G., B. Cooke, and S. Jennings. 2016. Estimating density-dependent impacts of
714 European rabbits on Australian tree and shrub populations. Australian Journal of
715 Botany **64**:142-152, doi:10.1071/bt15208.
716 Mutze, G., J. Kovaliski, K. Butler, L. Capucci, and S. McPhee. 2010. The effect of rabbit
717 population control programmes on the impact of rabbit haemorrhagic disease in

718 south-eastern Australia. Journal of Applied Ecology **47**:1137-1146,
719 doi:10.1111/j.1365-2664.2010.01844.x.
720 Noble, I. R. 1977. Long-Term Biomass Dynamics in an Arid Chenopod Shrub Community at
721 Koonamore, South Australia. Australian Journal of Botany **25**:639-653,
722 doi:10.1071/Bt9770639.
723 Pagel, J., B. J. Anderson, R. B. O'Hara, W. Cramer, R. Fox, F. Jeltsch, D. B. Roy, C. D.
724 Thomas, F. M. Schurr, and S. McMahon. 2014. Quantifying range-wide variation in
725 population trends from local abundance surveys and widespread opportunistic
726 occurrence records. Methods in Ecology and Evolution **5**:751-760, doi:10.1111/2041-
727 210x.12221.
728 Pech, R. P., and G. M. Hood. 1998. Foxes, rabbits, alternative prey and rabbit calicivirus
729 disease: consequences of a new biological control agent for an outbreaking species in
730 Australia. Journal of Applied Ecology **35**:434-453, doi:10.1046/j.1365-
731 2664.1998.00318.x.
732 Phillips, B. L., G. P. Brown, J. K. Webb, and R. Shine. 2006. Invasion and the evolution of
733 speed in toads. Nature **439**:803, doi:10.1038/439803a.
734 Plummer, M. 2016. rjags: Bayesian Graphical Models using MCMC.
735 Poole, W. E. 1960. Breeding of the of wild rabbit, *Oryctolagus cuniculus* (L.), in relation to
736 the environment. CSIRO Wildlife Research **5**:21-43.
737 R Core Team. 2017. R: A language and environment for statistical computing. R Foundation
738 for Statistical Computing, Vienna, Austria.
739 Rout, T. M., and T. Walshe. 2013. Accounting for Time Preference in Management
740 Decisions: An Application to Invasive Species. Journal of Multi-Criteria Decision
741 Analysis **20**:197-211, doi:10.1002/mcda.1490.
742 Roy-Dufresne, E., M. Lurgi, S. C. Brown, K. Wells, B. Cooke, G. Mutze, D. Peacock, P.
743 Cassey, D. Berman, B. W. Brook, S. Campbell, T. Cox, J. Daly, I. Dunk, P. Elsworth,
744 D. Fletcher, D. M. Forsyth, G. Hocking, J. Kovaliski, M. Leane, B. Low, M.
745 Kennedy, J. Matthews, S. McPhee, C. Mellin, T. Mooney, K. Moseby, J. Read, B. J.
746 Richardson, K. Schneider, E. Schwarz, R. Sinclair, T. Strive, F. Triulcio, P. West, F.

747 Saltre, and D. A. Fordham. 2019a. The Australian National Rabbit Database: 50 yr of
748 population monitoring of an invasive species. Ecology **100**:e02750,
749 doi:10.1002/ecy.2750.
750 Roy-Dufresne, E., F. Saltré, B. D. Cooke, C. Mellin, G. Mutze, T. Cox, and D. A. Fordham.
751 2019b. Modeling the distribution of a wide-ranging invasive species using the
752 sampling efforts of expert and citizen scientists. Ecology and Evolution **9**:11053-
753 11063, doi:10.1002/ece3.5609.
754 Royle, J. A. 2004. N-mixture models for estimating population size from spatially replicated
755 counts. Biometrics **60**:108-115, doi:10.1111/j.0006-341X.2004.00142.x.
756 Schurr, F. M., J. Pagel, J. S. Cabral, J. Groeneveld, O. Bykova, R. B. O'Hara, F. Hartig, W.
757 D. Kissling, H. P. Linder, G. F. Midgley, B. Schröder, A. Singer, and N. E.
758 Zimmermann. 2012. How to understand species' niches and range dynamics: a
759 demographic research agenda for biogeography. Journal of Biogeography **39**:2146-
760 2162, doi:10.1111/j.1365-2699.2012.02737.x.
761 Simberloff, D. 2003. How Much Information on Population Biology Is Needed to Manage
762 Introduced Species? Conservation Biology **17**:83-92, doi:10.1046/j.1523-
763 1739.2003.02028.x.
764 Sofaer, H. R., C. S. Jarnevich, and I. S. Pearse. 2018. The relationship between invader
765 abundance and impact. Ecosphere **9**:e02415, doi:10.1002/ecs2.2415.
766 Stodart, E., and I. Parer. 1988. Colonisation of Australia by the rabbit Oryctolagus cuniculus
767 (L.). Canberra, CSIRO.
768 Strayer, D. L., V. T. Eviner, J. M. Jeschke, and M. L. Pace. 2006. Understanding the long-
769 term effects of species invasions. Trends in Ecology & Evolution **21**:645-651,
770 doi:10.1016/j.tree.2006.07.007.
771 Strive, T., P. Elsworth, J. Liu, J. D. Wright, J. Kovaliski, and L. Capucci. 2013. The non-
772 pathogenic Australian rabbit calicivirus RCV-A1 provides temporal and partial cross
773 protection to lethal Rabbit Haemorrhagic Disease Virus infection which is not
774 dependent on antibody titres. Veterinary research **44**:51-51, doi:10.1186/1297-9716-
775 44-51.

776 Tabor, K., and J. W. Williams. 2010. Globally downscaled climate projections for assessing
777 the conservation impacts of climate change. Ecological Applications **20**:554-565,
778 doi:10.1890/09-0173.1.
779 Wells, K., P. Cassey, R. G. Sinclair, G. J. Mutze, D. E. Peacock, R. C. Lacy, B. D. Cooke, R.
780 B. O'Hara, B. W. Brook, and D. A. Fordham. 2016a. Targeting season and age for
781 optimizing control of invasive rabbits. Journal of Wildlife Management **80**:990-999,
782 doi:10.1002/jwmg.21093.
783 Wells, K., D. A. Fordham, B. W. Brook, P. Cassey, T. Cox, R. B. O'Hara, and N. I.
784 Schwensow. 2018. Disentangling synergistic disease dynamics: Implications for the
785 viral biocontrol of rabbits. Journal of Animal Ecology **87**:1418-1428,
786 doi:10.1111/1365-2656.12871.
787 Wells, K., R. B. O'Hara, B. D. Cooke, G. J. Mutze, T. A. Prowse, and D. A. Fordham. 2016b.
788 Environmental effects and individual body condition drive seasonal fecundity of
789 rabbits: identifying acute and lagged processes. Oecologia **181**:853-864,
790 doi:10.1007/s00442-016-3617-2.
791 Whittaker, R. J., M. B. Araújo, P. Jepson, R. J. Ladle, J. E. M. Watson, and K. J. Willis.
792 2005. Conservation Biogeography: assessment and prospect. Diversity and
793 Distributions **11**:3-23, doi:10.1111/j.1366-9516.2005.00143.x.
794 Williams, D., P. Acevedo, C. Gortázar, M. A. Escudero, J. L. Labarta, J. Marco, and R.
795 Villafuerte. 2007. Hunting for answers: rabbit (Oryctolagus cuniculus) population
796 trends in northeastern Spain. European Journal of Wildlife Research **53**:19-28,
797 doi:10.1007/s10344-006-0056-0.
798 Wilson, A. J., L. E. Kruuk, and D. W. Coltman. 2005. Ontogenetic patterns in heritable
799 variation for body size: using random regression models in a wild ungulate
800 population. American Naturalist **166**:E177-192, doi:10.1086/497441.
801 Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in
802 space and time. Trends in Ecology & Evolution **16**:446-453, doi:10.1016/S0169-
803 5347(01)02205-4.
804

Data Accessibility Statement:

All output maps (netcdf format) and R code to reproduce the hotspot analysis and Figure 6 is available in Data S1. Spotlight data used for the analysis is available in Roy-Dufresne et al. (2019a).

Tables:

Table 1: Coefficient estimates (± S.D) for the climatic covariates used to model detection probability of rabbits during spotlight surveys in different climate zones. Geographic locations of the climate zones can be found in Figure 2. Descriptions of the climate zones are provided in Appendix S1: Table S2.

Climate Zone	Daily maximum temperature	Daily minimum temperature	Daily Precipitation
1	-0.02 (0.10)	0.08 (0.11)	0.05 (0.05)
2	-0.34 (0.15)	0.04 (0.11)	0.03 (0.05)
3	0.52 (0.17)	0.00 (0.11)	0.07 (0.13)
4	-0.30 (0.21)	-0.02 (0.14)	0.04 (0.12)
5	0.02 (0.41)	0.03 (0.17)	0.05 (0.14)
6	-0.28 (0.23)	0.00 (0.10)	0.05 (0.09)
7	-0.20 (0.17)	0.01 (0.13)	0.01 (0.10)
8	0.26 (0.17)	0.09 (0.12)	0.08 (0.11)
9	0.30 (0.13)	0.07 (0.10)	0.07 (0.09)

Table 2: Performance of the random forest regression and classification models. Cross validated model performance was assessed using 1000 repeats of Monte-Carlo cross validation with an 80/20 (train/test) split. Accuracy of both the regression and classification models was excellent across climate zones and in most cases fair to excellent within climate zones (Hosmer et al. 2013). For climate zones 5 and 7, mean balanced accuracy and true skill statistics could not be calculated as the classification model predicted all samples as belonging to the same class.

	Classification					Regression		
	Sensitivity	Specificity	Mean Balanced Accuracy	True Skill Statistic		Mean observed	Mean predicted	Mean absolute error
Overall	0.86	0.85	0.86	0.71		6.4 (±11.9)	4.9 (± 6.4)	1.5
Climate Zone								
1	0.68	0.76	0.72	0.44		8.7 (±13.9)	6.2 (±6.6)	2.6
2	0.87	0.50	0.68	0.37		5.7 (±9.9)	4 (±5.8)	1.7
3	0.60	1.00	0.80	0.60		14.7 (±20.1)	10 (±8.2)	4.7
4	0.49	0.87	0.68	0.36		11.5 (±16.3)	8.6 (±6.4)	2.9
5	NA	1.00	NA	NA		7.4 (±7.3)	6.2 (±1.9)	1.2
6	0.67	0.90	0.78	0.57		5.3 (±6.9)	5.2 (±4.3)	0.1
7	1.00	NA	NA	NA		1.6 (±2.5)	1.4 (±1.1)	0.1
8	0.88	0.68	0.78	0.56		12.9 (±15.4)	11 (±11)	2.0
9	0.99	0.50	0.75	0.49		2.1 (±4.1)	1.6 (±1.4)	0.6

Figure legends:

Figure 1: The modelling approach used to correct records of abundance from spotlight surveys and produce maps of rabbit density, abundance, and hotspots. The hierarchical Bayesian N-Mixture model was created using abundance records from spotlight survey data and covariates linked to detection probability (a). The bias-corrected outputs of this model were then used to make spatiotemporal explicit predictions of rabbit density (b) and abundance (c). Note that the random forest abundance estimation (c) was dependent on the random forest density classification (b) as both a point based measure (i.e. for training the abundance model), and as a continuous spatial coverage (i.e. for predicting maps of estimated abundance). Items with a dashed outline represent point-based (i.e. individual grid-cell) data, solid outlines represent continuous spatial coverage (i.e. pixels over the entire grid). Arrows show the connections between the various elements.

Figure 2: Map showing the location of all spotlight rabbit abundance records (1971-2012) and the sites used in the final analysis. Column plots show the relative magnitudes for observed (dark grey) and Bayesian model corrected abundance (light grey) for a selection of sites. The different colors represent the nine climatic zones that were used in our analysis. Descriptions of the climate zones are in Appendix S1: Table S2. The beige color represents areas that are outside of the climatic envelope of our sampling sites. Map is projected in GDA 1994 Australian Albers (EPSG:3577).

Figure 3: Spatial predictions of rabbit densities and abundances inferred from the random forest models. (a) Spatial variation in the mode of predicted corrected densities between 1972 and 2012. (b) Mean corrected abundance (rabbits/km of transect) between 1973 and 2012. (c) The probability of a correct model classification (low/high), with darker colors (dark green/dark red) indicative of high confidence in the predicted class. (d) The standard deviation of the predicted corrected abundances between 1973 and 2012. All maps are masked to the climatic envelope of the climate zones shown in Figure 2. The muted colors

represent areas with low prediction confidence i.e., pixels where the model is forced to extrapolate beyond the climate and environmental envelope used to calibrate the model. Maps are projected in GDA 1994 Australian Albers (EPSG:3577).

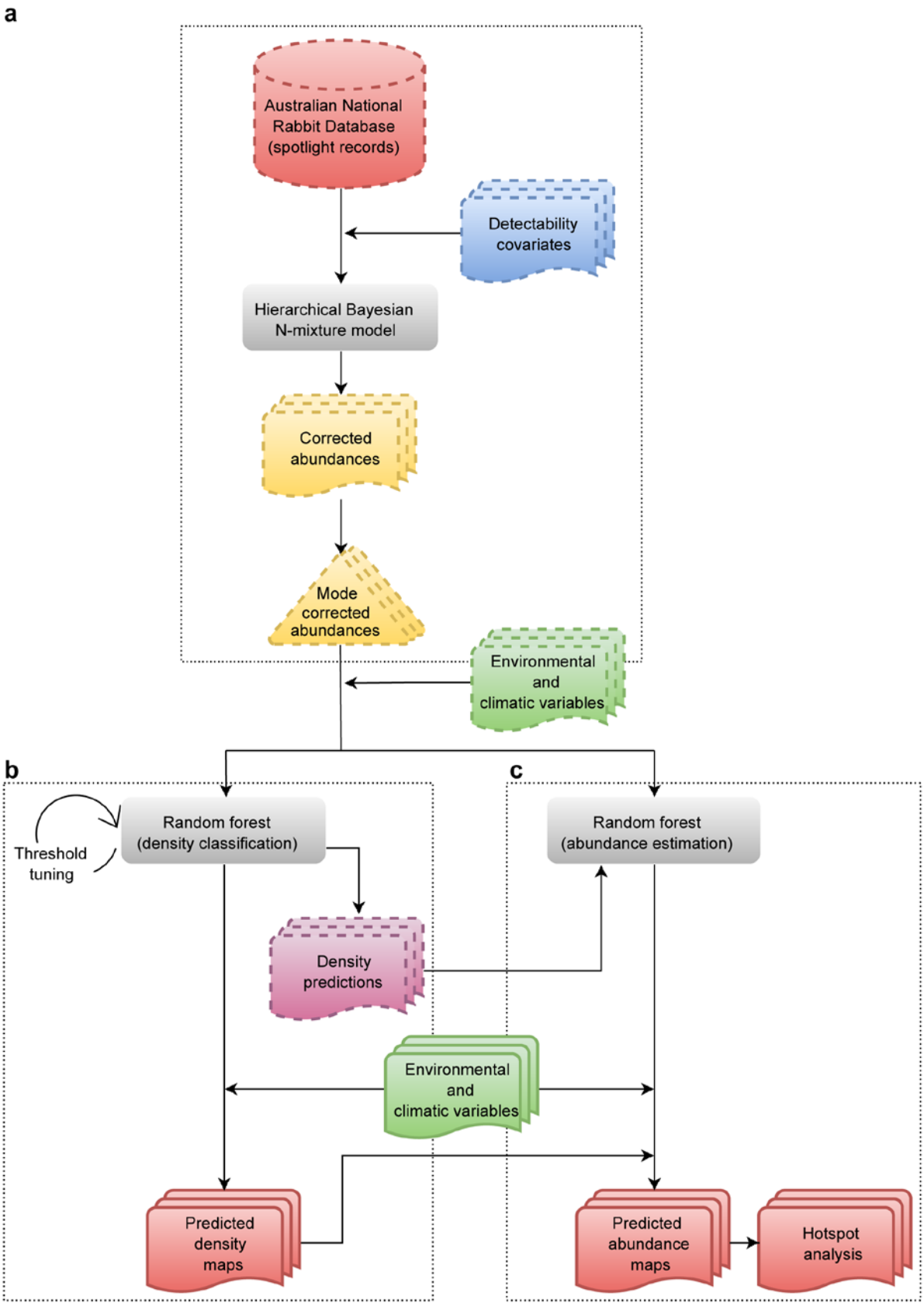
Figure 4: Accumulated Local Effects (ALE) plots of four important variables depicting the variation in rabbit abundance across environmental gradients. Values on the y-axes represent changes in abundance from the average sample conditional on the values on the x-axes (a). The rug plots on the inside of the x-axis show the distribution of sites across that variable. The 3-dimensional plots (b, c) show the second-order effects of benign RCV-A1 calicivirus with two climatic variables; average temperature two years prior to the survey (b), and total precipitation the year prior to the survey (c), after removing the main effects of the features. In other words, panels b and c only show the additional interaction effect of the two features. The colors correspond to the values on the z-axis, with reds indicating a more positive effect on abundances.

Figure 5: Empirically corrected abundance estimates (i.e. corrected; black line) and projected (blue line) abundance estimates of rabbits for nine selected climate zones in Australia. Points display the median estimates of corrected abundance estimates and predicted abundances (based on the random forest regression model) across all sites within a climate zone for a given year. The lines are for connecting points between sampling years, but do not represent the estimates between years, and show the discontinuous nature of the sampling within some climate zones. The blue bandings represent the 95% (light blue) and 50% (dark blue) prediction intervals for the regression model predictions within each zone.

Figure 6: Percent time spent in hotspot conditions for high rabbit abundances. Values for each cell (a) represent the proportion of time (between 1973 and 2012) that cells were considered “hotspots” of rabbit abundance. The map is masked to the climatic envelope of the climate zones shown in Figure 2. Muted colors represent areas with low prediction confidence i.e., pixels where the RF model was forced to extrapolate beyond the climate and

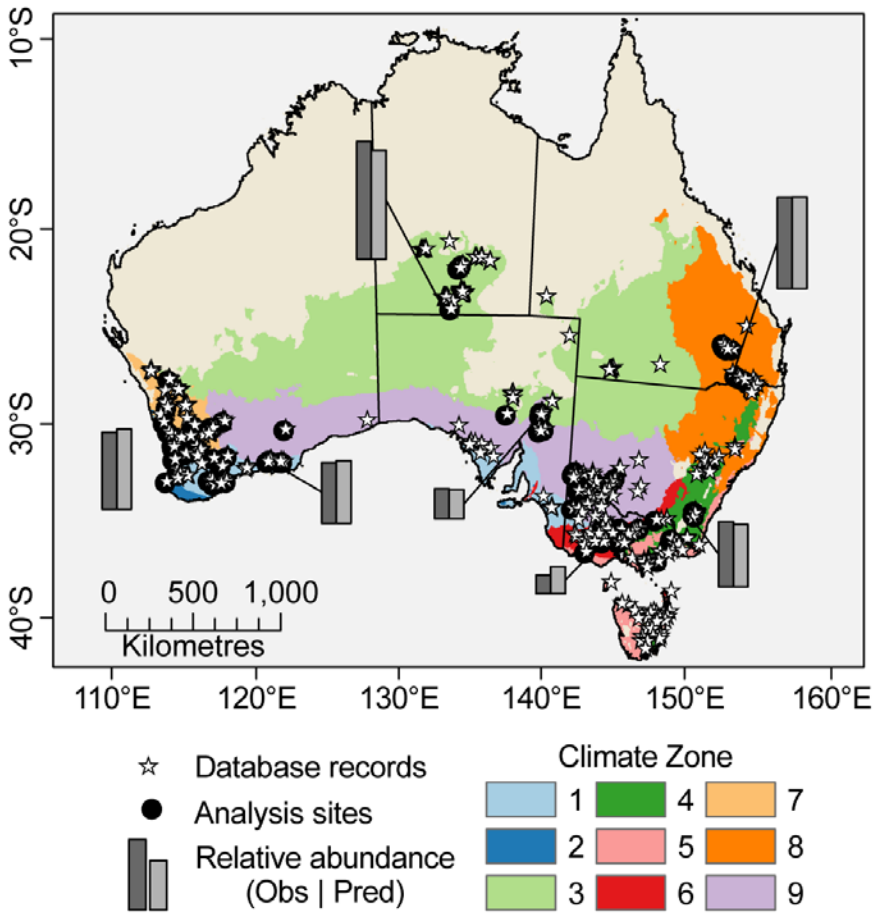
environmental envelope used to calibrate the model. Map is projected in GDA 1994 Australian Albers (EPSG:3577). The cumulative relative distribution frequency plots (b) show the inter-annual variability in rabbit populations for the RF predicted abundances that were used to define hotspot thresholds.

879 **Figures:**



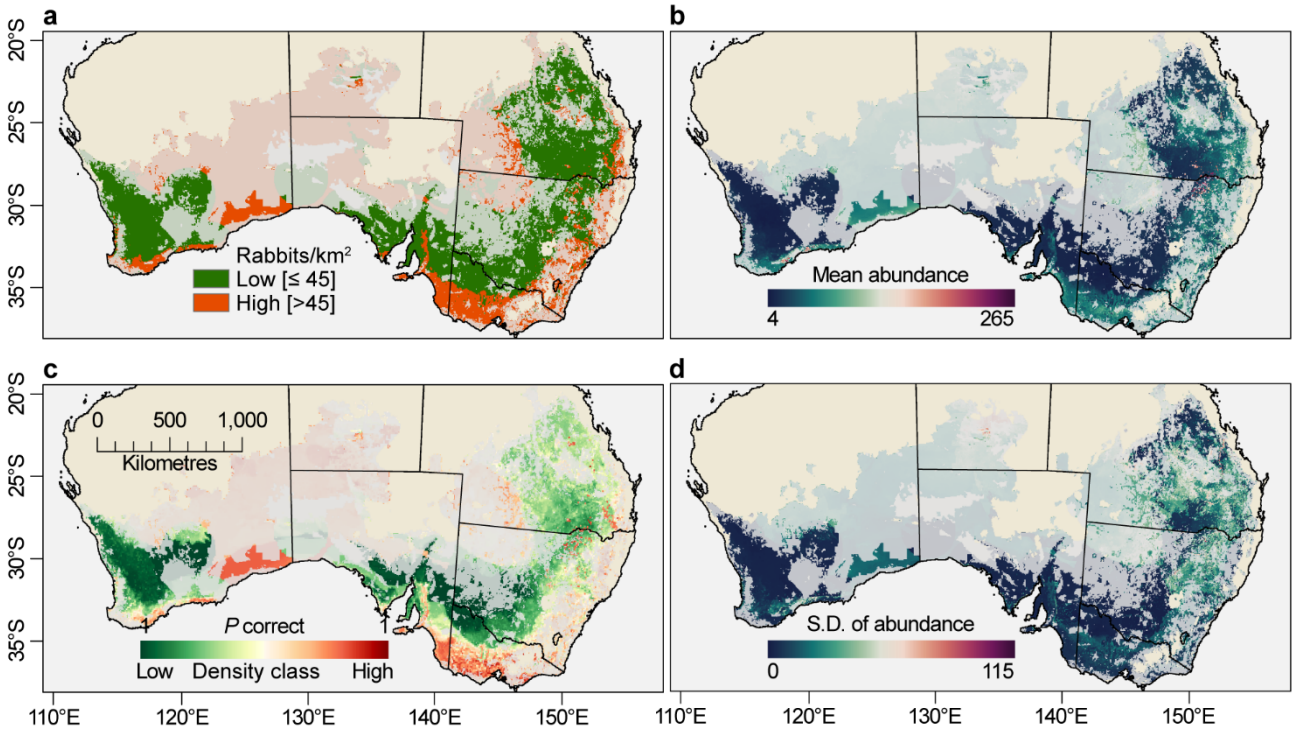
880

881 Figure 1



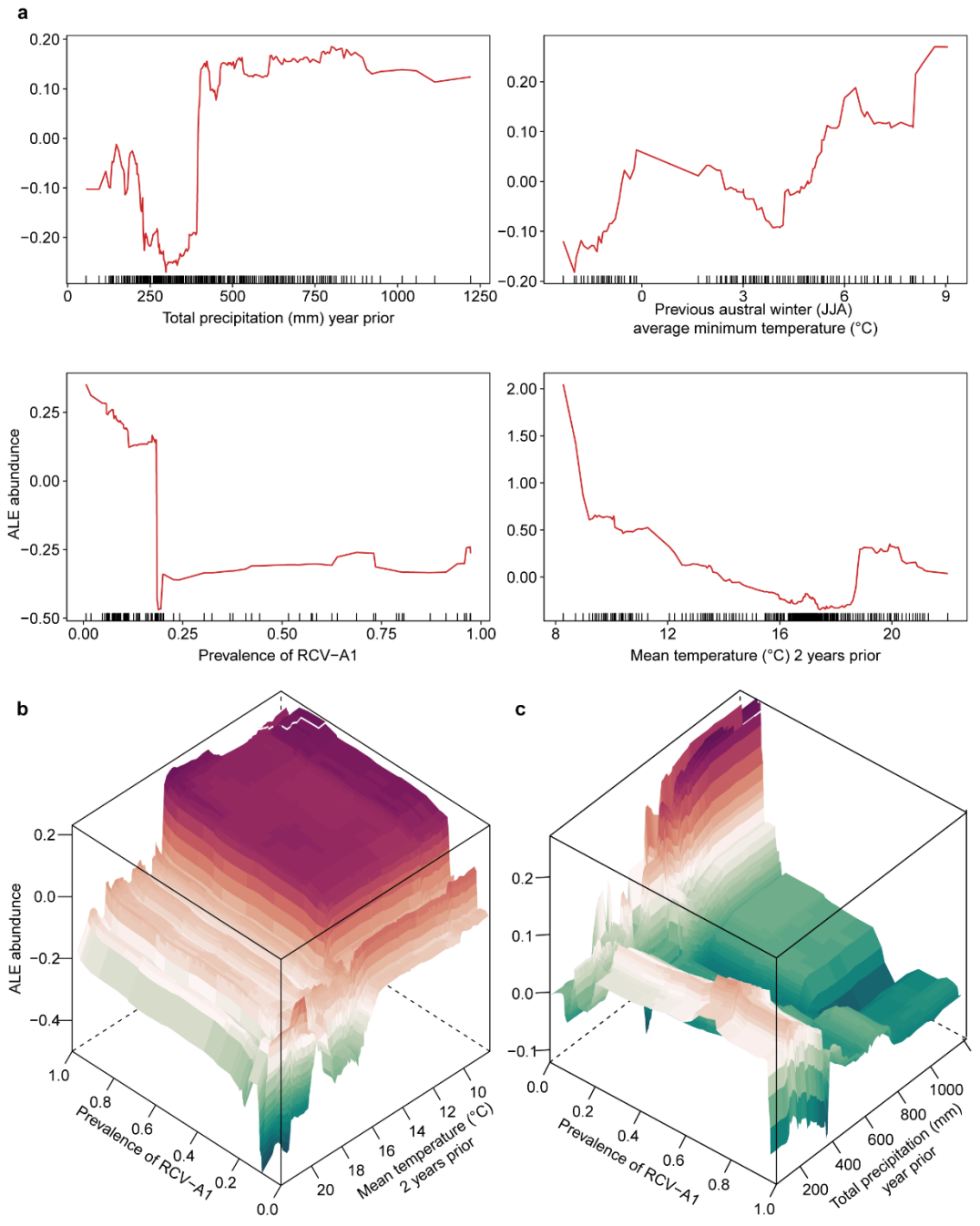
882

883 Figure 2



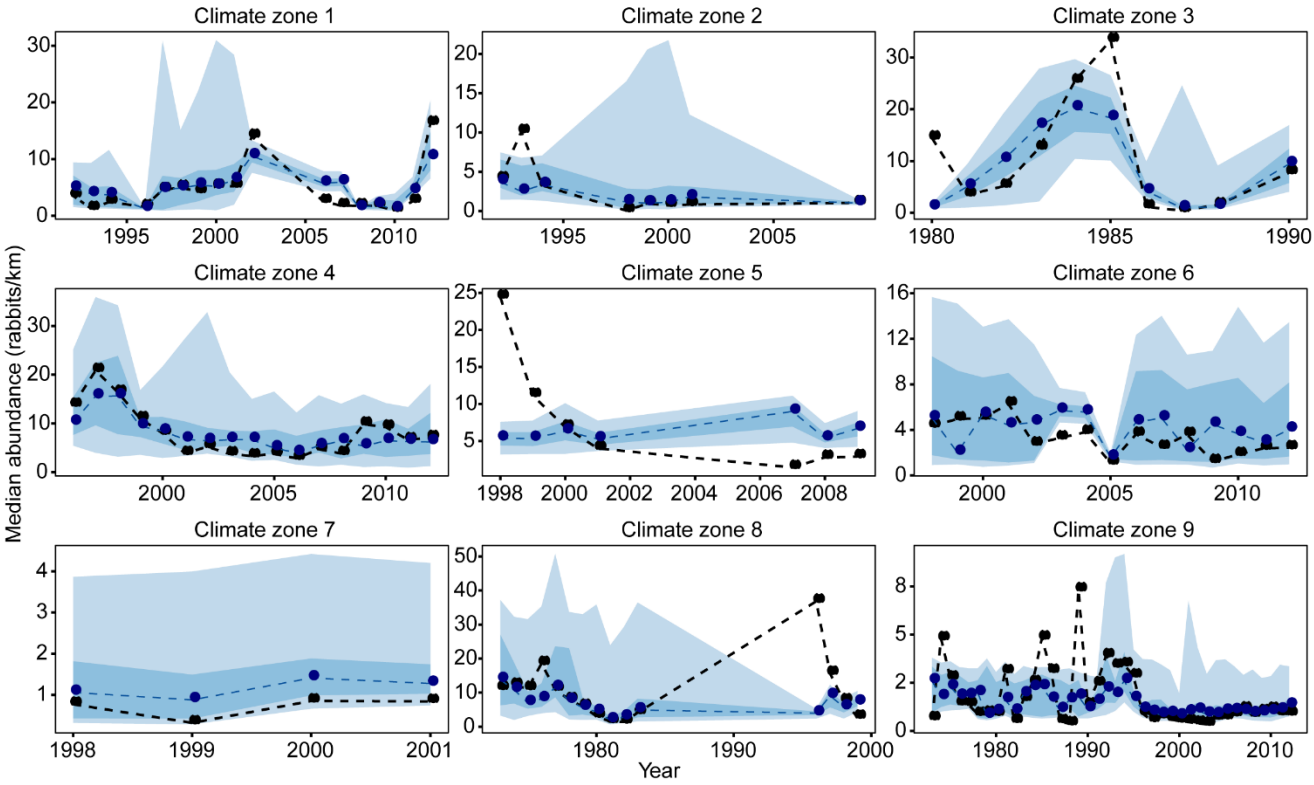
884

885 Figure 3



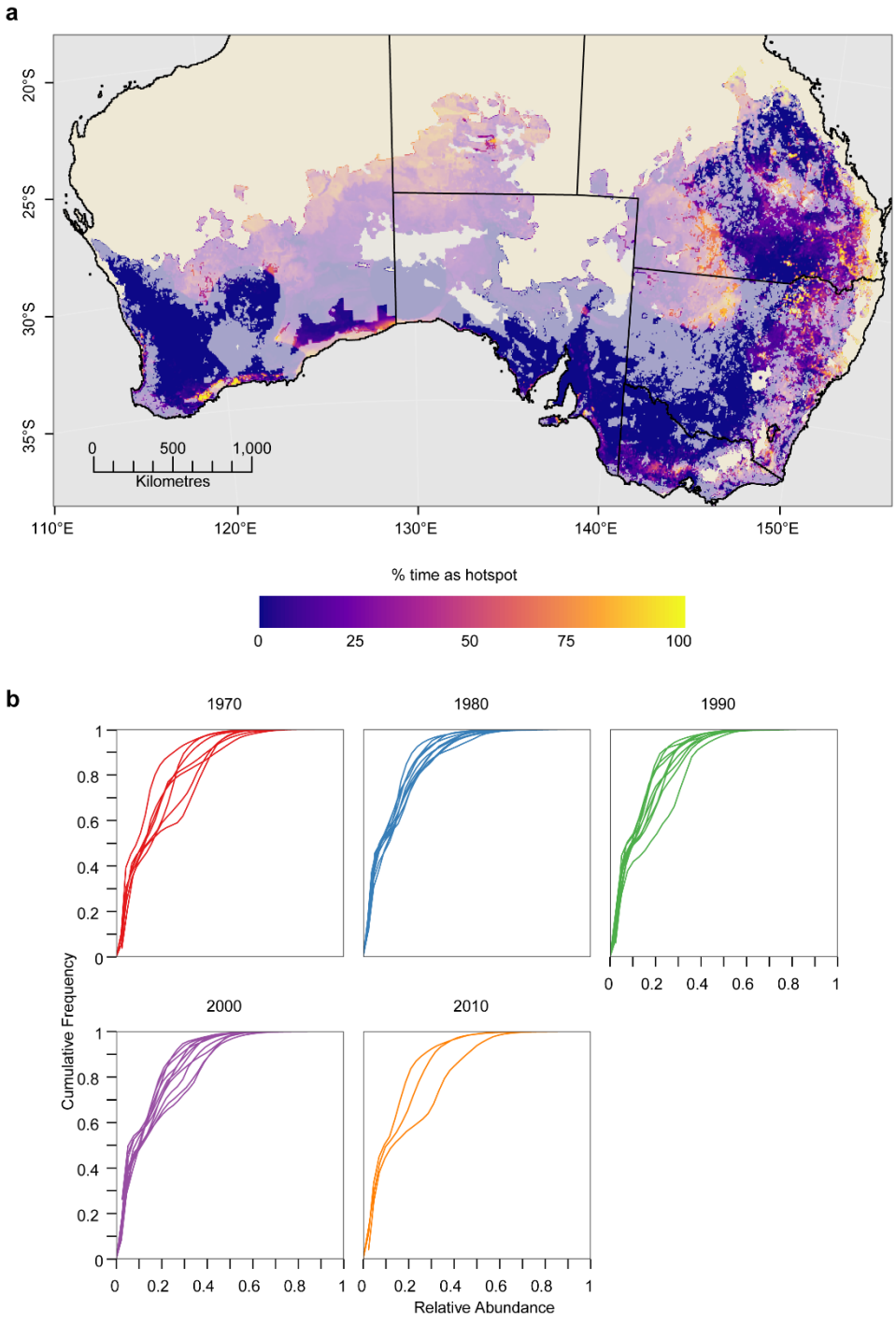
886

887 Figure 4



888

889 Figure 5



890

891 Figure 6