Australia's National
Science Agency

A preliminary national baseline population estimate for koalas

Testing the analytical framework for generating a national koala population estimate using historic observations and surveys

15 May 2023

Citation

CSIRO (2023). *A preliminary national baseline population estimate for koalas*. Progress Report. CSIRO

Table of Contents

1 Introduction

This report provides the background to the National Koala Monitoring Program's first annual reporting of the estimated national population of koalas. Contained in this report are contemporary estimates for both the listed (Queensland, New South Wales and the Australian Capital Territory) and the unlisted (Victoria and South Australia) populations, alongside mapping of the current estimated distributions for these areas. We detail the modelling and data assumptions used to derive this estimate, and give background to the approach and inputs used. The results of this study are being prepared into a peer reviewed scientific paper which will be made available online once published.

1.1 Background

In February 2022, the koala (combined populations of Queensland, New South Wales, and the Australian Capital Territory) was up-listed to 'Endangered' under the Environment Protection and Biodiversity Conservation Act 1999 (EPBC Act). Previous estimates have relied on more qualitative approaches such as expert elicitation to try and fill this gap.

The National Koala Monitoring Program (NKMP) provides a robust, data-driven approach to deriving koala population estimates across both the koala's listed (Qld, NSW, ACT) and unlisted (Vic and SA) ranges. The program will achieve this by designing and implementing an inclusive monitoring and modelling approach which enables the integration of multiple different sources of data and knowledge into processes which are established to ensure a long-lasting and robust monitoring program.

This report presents the results of an initial baseline assessment of koala populations which brings together the best available historic datasets to generate new estimates, and credible/confidence intervals, for both the listed and unlisted populations of koala. This analysis is undertaken even though the monitoring program is in its infancy and is still building its on-ground monitoring capacity. The challenges of 'limited' and 'fragmented' data which have, to date have made quantitative assessment of koala populations at the national level problematic still exist and are still limiting. However, the preliminary results presented in this report show the NKMP's capability to develop and implement new analytical approaches to integrate these fragmented and limited data to form nationally consistent estimates. These analytical approaches are expected to pair existing fragmented data with the new data being collected within the national monitoring program. The new data will form a backbone that the existing data cannot, but the existing data will give regional information that the national data cannot. The result is an integrated national program that will be able to provide improved results with greater levels of confidence.

1.2 Modelling rational and assumptions

The National Koala Monitoring Program's modelling approach was developed to enable the meaningful use of as many different data sources and monitoring methods as possible. This flexibility enables us to utilise data coming from multiple stakeholders and regional scale monitoring programs and allows for an inclusive approach to on-ground monitoring. This modelling approach has been designed to operate with a robust technical backbone of monitoring sites, arising from a nationally coordinated design that is still being implemented. The current baseline estimate relies on the use of available historic data sources. To enable this, we made a number of assumptions about these data including:

1) The last ten years of data (1 Jan 2013 to 31 Dec 2022) is meaningful and relevant to contemporary populations. Without this assumption the data become geographically patchy and require many more strong assumptions to enable integration.
2) We assume that the presence-only data is consistently biased. That is, the pattern in which humans observe and report koalas is consistent through different states and locations, and also consistent through time.
3) Covariates, described in Section 3, are used to describe the environmental conditions that koalas prefer and those which they avoid. We have also included some covariates to describe human search-effort biases that are related to the way koalas are observed by humans rather than to variation in koala abundance. We have modelled the data at the finest resolution that we could achieve repeatable/reliable results (250 m grid cells). This is the common approach in the distribution modelling literature.
4) It is common practice to `thin' the presence-only data. This practice prevents the unrealistic clustering of data. We hypothesise that such clustering is due to the human observation process, such as sharing of koala information on various types of media, which could create a localised increase of observations. We assume that such random clustering is unrelated to koala abundance, and that it cannot be captured by covariates related to consistent sampling effort bias. To counter this, thinning was performed so that we incorporated one observation per 250 m grid cell.
5) We make the extremely common and pragmatic assumption that the survey data (presence/absence or single/double count, see Section 3.2) was collected in locations that were chosen independently of whether there are known koalas present or not. This *preferential sampling,* if present, could provide substantial bias into the results from the model. One way to overcome this problem is to randomise the survey locations, like those locations chosen for the NKMP.
6) All statistical models contain assumptions, and this model is no different. Such assumptions include statistical distribution assumptions, anisotrophy and stationarity of the spatial process, functional forms of the covariate effects, and so on. Given the disparate sources of data, many of these assumptions are required and are also untestable. With a consistent and coordinated national monitoring program such assumptions will become less important and are also more testable.

1.3 Model type and structure

The analytical method used, to describe the distribution and abundance of koalas, is an integrated species distribution model (ISDM, see Isaac et al., 2020). These models attempt to leverage the desirable attributes of the different data types whilst simultaneously mitigating their respective weaknesses (e.g. Fletcher Jr et al., 2019). As an example, presence-only data is useful as it is broadly available and is numerous (strengths), but it also contains bias (from non-constant search-effort).

The model contains descriptors of the variation due to the koala's distribution and the environmental covariates (temperature, MODIS continuous vegetation fields, soil moisture, feed tree predictions), as well as inclusion probabilities and a spatially explicit term (see Section 3 for a description of data and their sources). It also contains terms to model the search bias in the presence-only (PO) data and accessibility (Weiss et al., 2018). Differences in survey methods (e.g. from visual transects or from scat sampling) between the data sources are allowed for by allowing survey artefacts into the model. The landscape considered in the analysis is reduced to just those cells that have *any* native trees observed (using remotely sensed data).

Once the model is fitted/trained to the data, predictions can be made using the model's predictive/posterior distribution for each 5km x 5km cell. The posterior distribution is summarised for each raster cell using the 2.5% and 97.5% percentile and the median. Maps of koala probability of presence can be generated using these cell-based predictions. Overall koala abundance can be estimated using the sum of the individual cell values.

The most challenging part of the analysis is to choose which survey should be used for the intercept in the prediction model. Its choice will directly affect the level of the predictions and also the abundance estimate. This is not just a matter of some methods being more effective in detecting koalas, rather it is about preferential sampling and about localised sampling. For example, if a sampling method is only used in localised areas of high koala density, then there is a confounding between method and area. If this intercept were then used for a national abundance estimate, then it is likely to grossly over-predict the number of koalas. Conversely, for methods used only in poor habitat. The most robust way to make this decision is to utilise the sampling method that is employed most extensively and has been done with (near) random survey design. In the current dataset, we consider this to be single-count visual transects that are extensively performed in multiple states. It is a single count method, implying that the observer detection probability cannot be accommodated. Nevertheless, it is the best option available presently. We note that we do not use the current double-count data, even though this data-type has the ability to estimate detectability. This is because the individual double-count datasets are localised and are often in locations with known high koala habitats. The previously mentioned bias negatively outweighs this benefit.

2 Results

The NKMP modelling approach is developed to enable the integration of all available data sources to provide the best possible, national scale, estimates of koala population and distribution. The results presented here highlight this approach using datasets collected over the past 10 years. As the national program is deployed we will be able to update these estimates of the current state of the population with increasing confidence. It is also expected that the impact of any significant events in recent years will become clearer as more contemporary data is collected and modelled.

2.1 Population estimates

The national baseline estimate presented in this report is delivered with wide confidence bounds, representing the relative uncertainty of generating a large-scale estimate from currently available data sources. This is not surprising as the datasets used to derive this estimate are varied and do not represent a high-quality sample of the possible places that koalas can be found across their range. The NKMP monitoring design has been developed to counter this challenge and, as such, our confidence in subsequent estimates will increase as the technical monitoring backbone of the program is deployed.

Our current best available estimate for koala population size at the national scale is between 287,830 and 628,010 individuals. When separated into the listed and unlisted populations we estimate between 117,050 and 244,440 individuals (Listed populations of NSW, ACT and QLD) and 170,780 and 383,570 individuals (Unlisted populations of VIC and SA).

These estimates are broadly in line with previous estimates of koala population at large scales, across the time period for which we have data available, which have tended to have wide confidence margins and are highly varied between different studies (Table 1).

Table 1. Previous published estimates of koala populations size

Year	2001	2012	2012	2014	2018	2020	2020	2021	2021
Source	**SA Gov[1]**	**SA Gov[1]**	**Adams–Hosking *et al*. (2016)[2]**	**IUCN[3]**	**AKF[4]**	**Heard & Ramsey (2020)[5]**	**NSW Gov[6]**	**AKF[4]**	**Federal CA[7] based on Adams–Hosking *et al*. (2016)**
Aus			331,000 (144k–605k)	300,000 (100k–500k)	45,000–82,000			32,000–58,000	92,184
Qld			79,000 (33k–153k)		10,090–19,150			6,455–12,085	
NSW			36,000 (14k–73k)		11,010–15,520		20,000 (15k–30k)	6,040–9,605	
Vic			183,000 (77k–327k)		14,280–27,640	Native: 413,000 Plantation: 47,000		11,950–23,080	
SA	Kangaroo Island: 27,000	Adelaide Hills & Mt Lofty Ranges: 114,000	33,000 (19k–51k)		Excl. KI: 10,355–19,840			Excl. KI: 7,615–13,150	

Sources of previous regional and national koala estimates

[1]SA Gov:The South Australian Koala Conservation and Management Strategyhttps://cdn.environment.sa.gov.au/environment/docs/koala-conservation-and-management-strategy-gen.pdf
[2]Adams-Hosking et al. (2016)https://onlinelibrary.wiley.com/doi/epdf/10.1111/ddi.12400
[3]IUCN:https://www.iucnredlist.org/species/16892/166496779
[4]AKF:Koala Population Estimates. https://www.savethekoala.com/wp-content/uploads/2021/09/KoalaEstimates2021.pdf
[5]Heard & Ramsey: Modelling Koala abundance across Victoria. https://www.wildlife.vic.gov.au/__data/assets/pdf_file/0022/512752/Heard-and-Ramsey_Koala_Popn_Assessment_FINAL.pdf
[6]NSW Gov:https://www.environment.nsw.gov.au/-/media/OEH/Corporate-Site/Documents/Animals-and-plants/Threatened-species/koala-strategy-2022-220075.pdf
[7]Federal Conservation advice:http://www.environment.gov.au/biodiversity/threatened/species/pubs/85104-conservation-advice-12022022.pdf*Note: Population estimates are based on Adams-Hosking et al. (2016) with a percent decline based on expert-elicitation values for bioregional population declines.*

2.2 Testing the sensitivity of population estimates to new monitoring data

The NKMP is being developed to provide a technical backbone which offers the most accurate estimates of koala population status and trends through an innovative data analysis approach. This method unifies all available data sources, maximizing their strengths and mitigating their weaknesses. For instance, citizen science data, though extensive, is biased and lacks sufficient information on population abundance. On the other hand, systematic survey data, while less comprehensive, provides more accurate population abundance data. Our method formally integrates these data for a more complete view.

For this approach to be able to provide the best possible outputs, we also need to pair it with an on-ground technical backbone of monitoring sites. The NKMP team and technical Community of Practice have worked to identify candidate sites for this technical monitoring backbone, and the NKMP with partners are now in the process of implementing it on the ground.

The current estimates rely on the aggregation of koala monitoring datasets made available to the NKMP. While these data are extensive, we acknowledge there are still numerous gaps in our understanding of koala populations. Future estimates will rely more heavily on the new monitoring data being collected by the program, across our technical monitoring sites and other monitoring partnerships. We fully expect that our population estimates will change as these data are gathered, not from koala populations changing but from greater availability of high quality monitoring data.

We tested the possible influence of this expectation by undertaking a sensitivity analysis. We supplemented the data used in the current model with a synthetic dataset. This dataset was constructed in the following ways;

1) Use the proposed network of technical backbone monitoring sites as the spatial representation of potential new koala monitoring data
2) Interrogate the Atlas of Living Australia to identify if, in the previous 10 years, a koala sighting has been recorded within 2 km of each proposed site
3) Build a presence-absence data set where a presence represents a sighting has been seen within 2 km of a proposed monitoring site and an absence represents no sighting of a koala within 2 km of a proposed monitoring site

This synthetic dataset contains the assumption that the patterns of koala presence/absence seen within the ALA database are representative of the true underlying pattern of koala presence/absence. With this assumption, this dataset gives us our current best approximation of the influence that new monitoring data will have on population estimates. Using this dataset within the current NKMP population and distribution adjusts the model for areas where there is little or no data, but where we strongly suspect there are no longer koalas present and generates a listed koala population estimate between 86,000 and 176,000 koalas.

The downward shift in the results suggest that, as better data begin to come into the NKMP from increased on-ground monitoring efforts, we will see our population estimates to trend towards the lower intervals of the current population estimate released with this report.

2.3 Distribution maps

The results shown in Figure 1 and Figure 2 represent our best estimate of the realised distribution of koalas, given suitable habitat within each spatial location. This distribution includes the climatic and landscape drivers of koala distribution. Maps are presented to represent the 'probability that at-least one koala is present within a cell' so can be interpreted as green tones meaning a likely presence of koalas, to peach tones meaning a low probability that koalas are present in this location. Importantly, in contrast to many of the more commonly presented distributional approaches we can deliver these estimates with associated upper and lower boundaries of confidence. Similar to our population estimates, the wide range of these confidence bounds highlights the difficulties in creating estimates from fragmented and limited data sources.

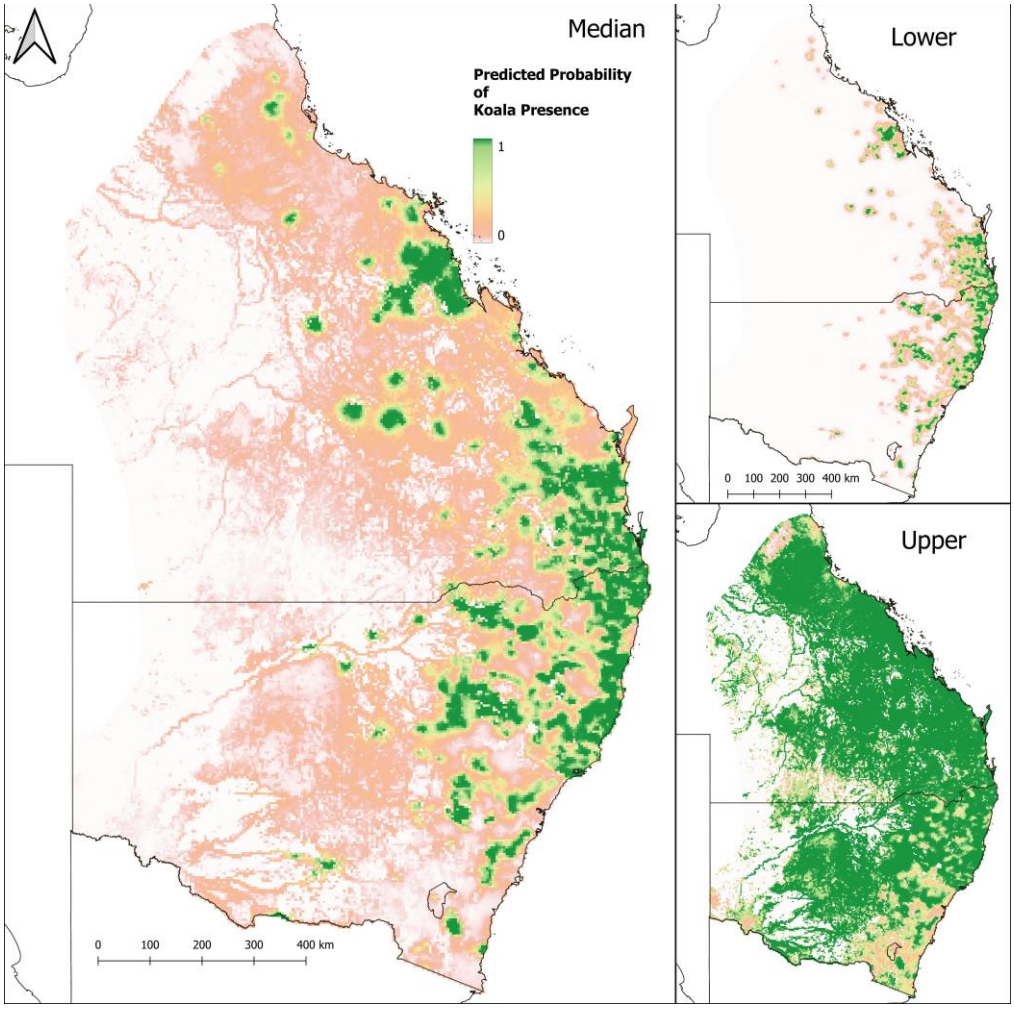

Figure 1. The predicted distribution of koalas across their listed range – QLD, NSW, ACT. Large panel on the left shows the median predictions whereas smaller panels show the upper and lower confidence estimates for the distribution. Grid cell size = 250 m.

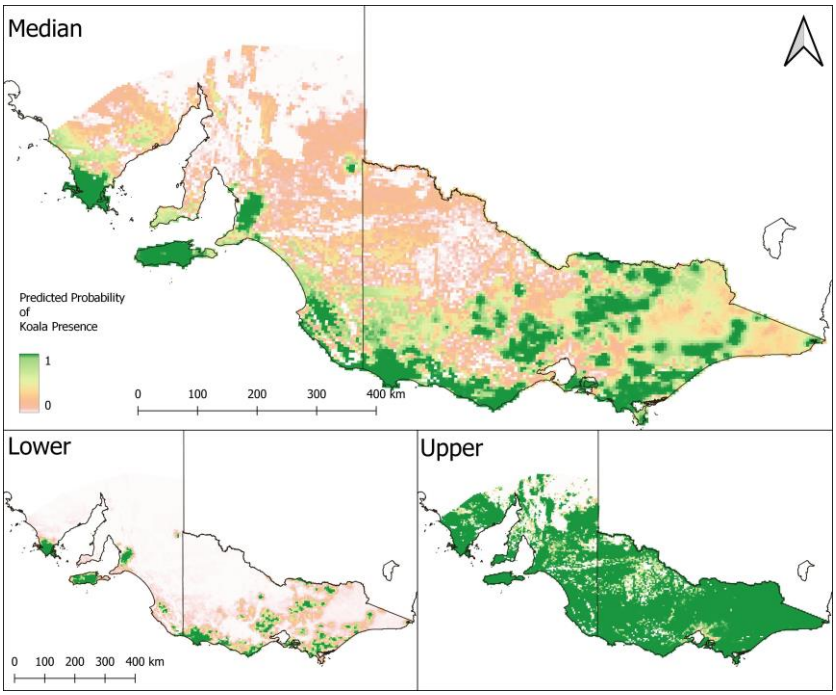

Figure 2. The predicted distribution of koalas across their unlisted range – VIC and SA. Large panel on the above shows the median predictions whereas smaller panels below show the upper and lower confidence estimates for the distribution. Grid cell size = 250 m.

3 Data inputs

3.1 Covariates and their purpose/assumptions

The national model has been constructed using the most parsimonious set of environmental and landscape feature variables available. This set was selected using an elicitation process at a national monitoring design workshop which brought together the NKMP Technical Community of Practice – a group consisting of 24 experts in koala science and ecology (Robinson et al., 2022) – to answer questions around the key national scale drivers of koala presence and abundance. From this workshop it was determined that, at the national scale, the key drivers of koala populations are;

- Thermal conditions: increased chance of heat stress and/or cold tolerance can limit koala populations.
- Water availability, particularly leaf moisture content: this drives a koala's ability to tolerate dry/drought conditions.
- The presence of suitable food trees: koalas have specific dietary requirements that should be met for populations to be sustained.

The NKMP team synthesised this advice into a set of covariates that represent these variables, given the best available national datasets. The results presented here rely on the following set of environmental and landscape variables.

Probability of a koala food tree being present: This is a custom developed mapping product which combines the species distributions of ~100 different tree species which have been identified as important food sources for koalas. Details on the development of this mapping product can be found in Hoskins et al. (2022). Data were scaled to have a mean of 0 and variance of 1 before being included into the model.

Maximum Temperature: This is a 30-year climatic average of the maximum temperature of the warmest month of the year (Harwood, 2019). Data were scaled to have a mean of 0 and a variance of 1 before being included into the model.

Soil Moisture content: The average soil moisture content across a period of 10 years. This is a remote sensing product developed by CSIRO and hosted by TERN (CSIRO & TERN, 2022). It is used as a proxy for water availability to koalas, both in leaves and as standing water. Data were skewed and so were log transformed to improve model fit, this was subsequently transformed into 2nd order orthogonal polynomials before being included into the model.

Density of woody vegetation: Taken from the MODIS continuous vegetation fields data product which predicts the proportion of woody vegetation contained within a 250m pixel (Dimiceli et al., 2015).Data were skewed and so were log transformed to improve model fit, this was subsequently transformed into 2nd order orthogonal polynomials before being included into the model. We scaled the polynomial components so that they had mean zero and variance 1.

Monitoring inclusions probability data product: One of the outputs of a NKMP workshop. This represents the probability that each cell should be included into the survey design. It is based on the environmental gradients that were identified as important at a national scale (Temperature, Moisture and Feed Trees). In the model it can be considered as an informed interaction of those variables.

Spatial positioning index along the Great Dividing Range: A custom derived index which provides a relative position along the length of the Great Dividing Range, starting at 1 in Northern Australia and increasing in value to South Australia. This is included to enable spatial variation in responses to environmental covariates between different parts of the koala's range. Data were scaled with a mean of 0 and variance of 1 before being included in the model.

Accessibility: Developed following Weiss et al. (2018)and represents the expected travel time from any place in the analysis area to the nearest population centre > 100,000 people. Included as a bias variable to account for the increased probability of presence-only data records of koalas closer to human population centres. Data are skewed and so were square-root transformed before being scaled to have a mean of 0 and variance of 1.

Probability of searching for mammals: A custom-derived index for presence-only datasets, approximating the increased koala observation rate within areas that are more intensively searched. This is approximated by taking a probability density of all Australian mammal records found within the Atlas of Living Australia.

Proportion of available habitat at a 5 km scale: Used to approximate landscape-scale processes, alongside local-scale processes captured above. Represents the proportion of a cell with possible koala habitat (see below).

The landscape considered in the analysis was masked to include only those areas which provided possible koala habitat, following consultation with experts about habitat features that determine the presence/absence of koalas. On that basis, we estimated possible koala presence/abundance only for those grid cells, at a 5 km^2 resolution, containing at least some native vegetation and including at least one tree that is 2 m or more in height. If a grid cell did not meet these conditions, it was excluded from the analysis.

3.2 Survey and observation datasets

Data representing a mixture of systematic surveys and presence-only observations were collated from multiple public and proprietary sources. These data represent a variety of different approaches to detect and quantify populations of koalas and make up four different classes of data which are integrated into the national model. These data classes are:

- **Presence-only observations** represent an observation of a koala that has been recorded and pushed into a publicly available database (e.g. the Atlas of Living Australia). These data provide no information on the search effort associated with an observation or any information about where koalas are absent.
- **Presence/absence observations** are usually collected during a systematic survey using methods and technology that provide an indication of the presence or absence of koalas. These data come from methods such as acoustic recorders, koala detection dogs and

systematic scat surveys. Such observations are linked to the survey effort undertaken, but they provide no information on the abundance of koalas (i.e. the number of koalas in an area).

- **Abundance/absence observations** come from systematic surveys which can provide an indication of the absence and density of koalas. These data come from systematic line transect surveys, drone surveys and a special class of transect survey called a **double-count** survey which provides additional information describing the probability of a surveyor detecting a koala, if it was present.

The data included within this baseline estimate come from 24 different data sources (not including presence-only data collected from the Atlas of Living Australia) across the four different data classes (Table 2). In total they represent 88,814 individual observations or surveys of koalas across a 10-year period. Data type and survey method were distributed non-randomly across the koala's range (Figure 3). This has limited our ability to confidently project population estimates across the complete range of the koala.

Table 2. Total numbers of records for the different data types and survey methods used in the NKMP baseline estimate

Data class	Survey method	Total records
presence-only	Observation	79,962
presence-absence	Acoustic recording	123
presence-absence	Diurnal koala search	121
presence-absence	Koala radial search	137
presence-absence	Koala SAT survey	1,262
presence-absence	Scat detector dog	4,062
single-count	Drone	43
single-count	Visual transect	1,417
double-count	Visual transect	1,687

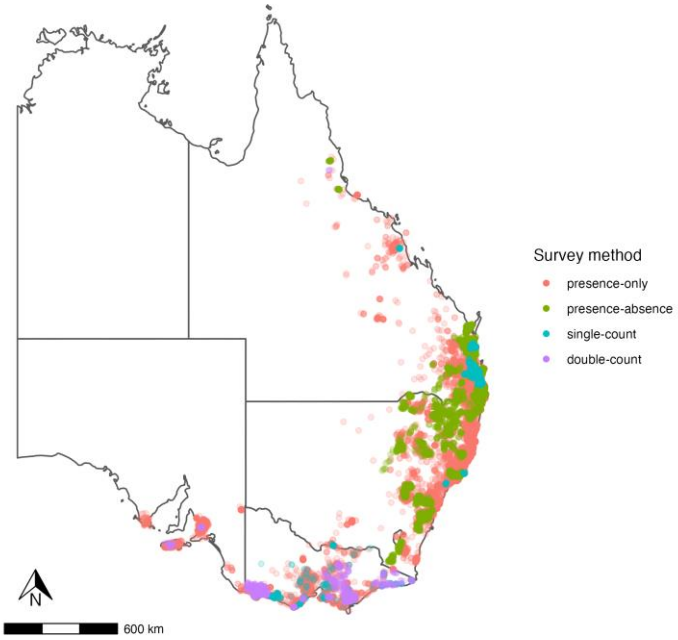

Figure 3. Map of the distribution of different data types across the range of the koala

4 Conclusion

This preliminary national baseline population estimate for koalas will continue to improve with better data and with evolving analytical methods to combine these existing data. The NKMP team are now working with several government agencies, universities, First Nation and local community groups, and non-government organisations to survey sites across the NKMP monitoring network. Different monitoring methods are incorporated into the NKMP model. This makes analysis complicated, but it also enables everyone to contribute to this national effort to deliver robust estimates and build a long-lasting capability to monitor and assess trends in koala populations.

5 References

CSIRO, & TERN. (2022). *The Soil Moisture Information Processing System (SMIPS)*. https://shiny.esoil.io/SMIPS/

Dimiceli, C., Carroll, M., Sohlberg, R., Kim, D., Kelly, M., & Townshend, J. (2015). MOD44B MODIS/terra vegetation continuous fields yearly L3 global 250m SIN grid V006. *NASA EOSDIS Land Processes Distributed Active Archive Center*, *10*.

Fletcher Jr, R. J., Hefley, T. J., Robertson, E. P., Zuckerberg, B., McCleery, R. A., & Dorazio, R. M. (2019). A practical guide for combining data to model species distributions. *Ecology*, *100*(6), e02710.

Harwood, T. (2019). *9s climatology for continental Australia 1976-2005: BIOCLIM variable suite* Version v1). https://doi.org/https://doi.org/10.25919/5dce30cad79a8

Hoskins, A. J., Robinson, C. J., McEvoy, J., Macdonald, S. M., Vanderduys, E., McKeown, A., Murray, J., Murphy, H., Marano, J., Golchin, M., Lourie, H. J., Henderson, A., Behzadina, S., Lloyd-Jones, L., & Foster, S. D. (2022). *The National Koala (Phascolarctos cinereus) Monitoring Program - Final Report*.

Isaac, N. J., Jarzyna, M. A., Keil, P., Dambly, L. I., Boersch-Supan, P. H., Browning, E., Freeman, S. N., Golding, N., Guillera-Arroita, G., & Henrys, P. A. (2020). Data integration for large-scale models of species distributions. *Trends in ecology & evolution*, *35*(1), 56-67.

Robinson, C. J., Murray, J., Hoskins, A. J., Foster, S. D., Vanderduys, E., Brown, S., Burke Da Silva, K., Cappie-Wood, T., Cristescu, R., Crowther, M., Gillespie, G., Hamilton, G., Hillyard, K., Rhodes, J., Law, B., Lawler, I., Lunney, D., Marsh, K., Martin, R., . . . Whisson, D. (2022). *Co-designing a national koala monitoring program. Summary of national koala monitoring program technical community of practice may workshop*.

Weiss, D. J., Nelson, A., Gibson, H. S., Temperley, W., Peedell, S., Lieber, A., Hancher, M., Poyart, E., Belchior, S., Fullman, N., Mappin, B., Dalrymple, U., Rozier, J., Lucas, T. C. D., Howes, R. E., Tusting, L. S., Kang, S. Y., Cameron, E., Bisanzio, D., . . . Gething, P. W. (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature*, *553*(7688), 333-336. https://doi.org/10.1038/nature25181

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiro.au/contact
csiro.au