Received: 25 July 2024 | Accepted: 13 January 2025
DOI: 10.1002/2688-8319.70018

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

RESEARCH ARTICLE

Supplementing human observation with artificial intelligence impacts demographic estimates for a critically endangered lizard

Emily A. Jordan[1,2] | John Ewen[2] | Rod Hitchmough[3] | Les R. Moran[3] | David J. Murrell[1] | J. Marcus Rowcliffe[2] | Lynn K. Adams[3]

[1]Department of Genetics, Evolution and Environment, Centre for Biodiversity and Environment Research, University College London, London, UK

[2]Zoological Society of London, London, UK

[3]Department of Conservation, Wellington, New Zealand

Correspondence
Emily A. Jordan
Email: emily.jordan@ioz.ac.uk

Funding information
Natural Environment Research Council, Grant/Award Number: DTP2 NE/S007229/1; Department of Conservation, New Zealand

Handling Editor: Gideon Deme

Abstract

1. Species conservation relies heavily on population estimates derived from capture–recapture analyses, which are liable to produce biased results if individual animals are incorrectly identified. Captive and known-animal studies have shown that supplementing human observation with artificial intelligence (AI) has the potential to reduce these errors. However, no study has directly quantified the relationship between using AI for individual identification and the demographic estimates it produces for a threatened population in situ.
2. We compared the demographic estimates produced by capture–recapture analyses of two distinct encounter histories constructed from the same survey data; one produced using individual identifications made by human observers alone (the 'human-only data set'), and one produced using AI software to aid individual identification (the 'AI-supplemented data set'). This approach enabled us to address two key questions: (i) does the use of artificial intelligence software for individual identification influence demographic estimates for an in situ conservation programme? (ii) How has the population of our case study species, the critically endangered Kapitia skink, responded following an extreme weather event, cyclone Fehi?
3. We found that, without AI, human observers appeared prone to make reclassification or 'splitting' errors, in which a recaptured animal was wrongly assigned as a new individual. Analysis of the AI-supplemented data set consistently produced lower estimates of population abundance over time, relative to the same analysis of the human-only data set. This provides new evidence that wild species monitoring efforts may be prone to underestimating the extinction risk of populations if they are dependent on individual identification methodologies with high potential for human errors.
4. Our case study species, the Kapitia skink, demonstrated a positive population trend in the period following cyclone Fehi. While promising, conservation intervention is recommended to address persistent threats.

5. *Practical implication*. Supplementing human observation with AI software for individual identification could mitigate errors leading to the underestimation of extinction risk for endangered species. We encourage further development of AI software to increase its automation and accessibility and recommend that practitioners consider its use in population monitoring based on the identification of individuals in imagery.

KEYWORDS
artificial intelligence, capture–recapture, individual identification, Kapitia skink, *Oligosoma salmo*, population estimation, reptile conservation, robust design

1 | INTRODUCTION

The identification of individual animals is fundamental in capture–recapture modelling. For species that display consistent individually unique patterning, photographic records of captured animals enable human observers to distinguish between individuals while bypassing the practical and welfare issues associated with traditional marking methods (Belaud et al., 2022; Petso et al., 2022; Vidal et al., 2021). However, identifying individual animals by the human eye alone is time-consuming and prone to error; and simulation studies have demonstrated that errors in individual capture histories due to misidentification could significantly impact the reliability of demographic estimates (Cruickshank & Schmidt, 2017; Bohnett et al., 2023; Johansson et al., 2020; Morrison et al., 2011; Strampelli et al., 2022). This impact may be particularly significant for rare, elusive species; estimate uncertainty is already high if elusivity results in low recapture rates. One solution is to account for misidentification in model formulation, though complex models may be impractical for everyday conservation needs (Morrison et al., 2011; Tucker et al., 2019; Yoshizaki et al., 2009). As robust estimates of demographic parameters are key for conservation planning, limiting sources of error is crucial, and research should promote user-friendly tools that increase confidence in parameter estimation.

Several studies have reviewed the use of artificial intelligence (AI) software to identify individual animals. AI can be trained to compare patterns between images and highlight similarities, which may be presented in user-friendly packages such as HotSpotter, APHIS or WildID (Bolger et al., 2012; Crall et al., 2013; Moya et al., 2015; Schneider et al., 2019). Outcomes are promising, with researchers reporting greater resource efficiency and accuracy in trials using AI to correctly classify images of known individuals (Bardier et al., 2020; Cruickshank & Schmidt, 2017; Dawson et al., 2021; Dunbar et al., 2021; Nipko et al., 2020; Park et al., 2019; Renet et al., 2019). Notably, Bohnett et al. (2023) reported that using AI software decreased reclassification errors, whereby observers mistakenly report a recaptured animal as a new individual. As this type of error is likely to lead to an overestimation of abundance, it has the potential to underestimate the extinction risk of threatened populations, suggesting a need for caution in interpreting demographic estimates for conservation decision-making and for research to quantify these effects in situ (Johansson et al., 2020; Morrison et al., 2011; Yoshizaki et al., 2009). Here, we present a case study in the context of active conservation planning for a highly threatened reptile.

The Kapitia skink, *Oligosoma salmo* (Melzer et al., 2019), is a small lizard endemic to Aotearoa New Zealand. Like many skinks in the genus, *O. salmo* is viviparous and omnivorous, but is unusual in having a prehensile tail (van Winkel et al., 2018). Following historic habitat loss and modern land conversion to pasture, the species' only wild population persists within a narrow area of non-native coastal habitat less than 2 km^2, warranting an IUCN Red List status of critically endangered (Hitchmough, 2021). The species continues to be threatened by invasive predators, coastal erosion and extreme weather events (Nelson et al., 2016; van Winkel et al., 2018). In 2018, cyclone Fehi destroyed over half the known contemporary habitat, and the occurrence of extreme weather is forecast to increase with changing climates (Fitzharris, 2007; Safaei Pirooz et al., 2019). As such, urgent action is required to prevent the species' extinction, and an initial understanding of its demographics is desired to advise conservation management.

This study aimed to produce initial estimates of population abundance and survival probability for the critically endangered Kapitia skink, and to quantify the impacts of using AI image recognition software on these estimates. We hypothesised that supplementing human observation with AI software would (a) reduce the rate of individual misidentification errors, and therefore (b) reduce bias in our population estimates.

2 | MATERIALS AND METHODS

2.1 | Study area

The study site encompasses the extant range of the Kapitia skink on the West Coast of South Island, Aotearoa New Zealand (42°36′ 57″ S, 171°5′2″ W), a narrow strip of non-native grass and scrubby vegetation circa 1 km long and 5 m wide. It is flanked by a sandy beach and close-cropped pastureland to the west and east, respectively, with close-cropped grass and a stream delineating the north and south edges. A linear array of Artificial Cover Objects (ACOs) made of crenulated Onduline™ (30 × 30 cm) spaced, on average, 8 m apart was used to attract skinks for capture.

FIGURE 1 Examples of catalogued images of distinct *Oligosoma salmo* individuals demonstrating the variation in scale patterning.

2.2 | Capture surveys

Data collection followed a robust design mark–recapture protocol in which surveys were divided into primary and secondary sampling occasions. In a robust design structure, a population is assumed to be closed to gains or losses within a primary occasion, then open to population change between primary occasions. Surveys were conducted between October and April, corresponding to warmer temperatures in the austral summer when the species is most active. Five primary sampling periods were completed within a 3-year span (2019–2021) following cyclone Fehi in 2018. Each primary occasion spanned 12 days, allowing for poor weather but maintaining a reasonable assumption of population closure for the species, and the time between primary occasions varied from 2 to 10 months depending on seasonality and surveyor availability. Either 10 or 11 surveys (secondary occasions) were conducted within each primary sampling period.

Surveys lasted on average 1.5 h and commenced between 09:00 and 13:00, having allowed the sun to warm ACOs to optimise the probability of skink captures. During each survey, every ACO was checked for skink presence. To aid capture, a bottomless plastic box was placed around the ACO before it was lifted, constraining any skinks present. Once captured, skinks were measured and photographed before being returned to their original location. Surveyors aimed to take consistent images of each skink; images of the left side were used for identification. Focus on a single side of each animal mitigated cataloguing and time constraints, while still utilising sufficient distinguishing features for identification. Identification relied on the unique individual scale patterning of lizards (Figure 1).

2.3 | Image processing

2.3.1 | Human-only

Images were sorted into a physical catalogue in which distinct individuals were identified and assigned unique numbers. Each image taken at a capture event was compared to this catalogue by four experienced surveyors (LA; RH; LM; and one other) and designated as either a recapture or a new individual. Records of individual captures were then constructed into encounter histories, with 1 s or 0 s, respectively, denoting the presence or absence of an individual during a sampling event. This collection of encounter histories, established using individual identifications made by human observation alone, is hereafter referred to as the 'human-only' data set.

2.3.2 | AI-supplemented

Digital images were then compared using 'HotSpotter' image recognition software (Crall et al., 2013) by an independent researcher (EJ). For this project, we favoured HotSpotter due to its reported high levels of accuracy (de Lorm et al., 2023; Dunbar et al., 2021; Monnet et al., 2022; Nipko et al., 2020) and ease of access (free at time of download). HotSpotter uses a feature-based approach, comparable to other popular wildlife recognition software including I^{3}S-Pattern (van Tienhoven et al., 2007) and WILD-ID (Bolger et al., 2012), which extracts descriptors of key areas of animal markings to infer matches. The focus on unique markings, as opposed to a pixel-by-pixel comparison, enables greater flexibility when photographing wildlife under varying conditions; it is more robust in handling differences in factors such as lighting, scale and direction (de Lorm et al., 2023).

To run HotSpotter, each image was cropped to a rectangular 'region of interest', the left side of each skink between its fore- and hind limbs, which together formed a database for image comparison. A 'query' was run on every image individually, inducing HotSpotter to compare that image with the full database and calculate scores indicating the level of similarity between images. In contrast to related software packages, HotSpotter runs two algorithms, *one*-vs-*one* and *one*-vs-*many*. In *one*-vs-*one*, the queried image is compared to each image in the database individually using the 'SIFT' method (Lowe, 2004), whereas in the *one*-vs-*many* algorithm each descriptor of the queried image is compared to every descriptor in the database using a nearest neighbour approach as in McCann and Lowe (2012). The similarity scores assigned to each individual are produced from the aggregated scores generated by both of these algorithms.

Following each run, the six highest scoring images were presented adjacent to the queried image and reviewed by the researcher. 'Hotspots' of matching features between images were highlighted to aid comparison and the researcher decided whether compared images were from the same individual. A second encounter history data set was produced using the identifications of individual captures made with the aid of HotSpotter, hereafter referred to as the 'AI-Supplemented' data set.

2.3.3 | Verification

Studies testing photographic methods for individual identification typically rely upon using tagged or otherwise previously known individuals to verify the accuracy of matches. In this case, to create a benchmark against which to verify the human-only and AI-supplemented data sets, we subjected all images on which human-only and AI-supplemented methods did not agree to additional inspection. This involved a further round of human scrutiny by the independent researcher, including with the aid of comparison tools in Hotspotter. The outcome of each image review was then discussed, and the accurate match was agreed upon between the experienced surveyors and the independent researcher. While we cannot rule out some remaining mismatching in the outcome of this process, in no case was this felt to be a serious risk. We refer to this benchmark as the Verified data set.

2.3.4 | Evaluating performance

To assess the relative performance of the image processing methods, we calculated their false acceptance rates (FAR), the frequency at which images of distinct individuals are recorded as matching pairs, and false rejection rates (FRR), the frequency at which images of the same individual are not matched (Jain, 2007), following Bardier et al. (2020):

$$\mathrm{FAR} = \frac{\text{Number of false matches}}{\text{Number of non-matching pairs}}.$$

$$\mathrm{FRR} = \frac{\text{Number of falsely rejected matches}}{\text{Number of true-matching pairs}}.$$

The number of non-matching pairs pertains to every possible comparison between images in which the photographed individuals were discrete and was determined by the number of possible comparisons, given by $n!/2\times(n-2)!$ minus the number of true-matching pairs, where n is the number of images compared.

2.4 | Capture–recapture analyses

Capture–recapture analyses were conducted in Program MARK using a robust design model, with a Huggins abundance estimator to derive estimates of population size N (Huggins, 1989, 1991; White & Burnham, 1999). The combination of open and closed periods in a robust design model enables the simultaneous estimation of survival and abundance, both important target parameters for conservation planning, as well as the inclusion of a temporary emigration parameter. Including temporary emigration in capture–recapture analysis accounts for the probability that some individuals may be unavailable for capture at a sampling occasion, which otherwise might violate the model assumption of equal recapture probability across individuals and bias estimates (Barker & White, 2004; Bird et al., 2014; Hammond, 1990). For the Kapitia skink, it is unlikely that individuals leave and return to the survey area for elongated periods; the land surrounding their remnant habitat is unsuitable for a species that actively prefers cover, and anecdotally, individuals show limited capacity for dispersal. However, lizards are known to spend long periods in natural refuges alternative to ACOs, which would exclude them from the surveyed population (Black et al., 2019; Lettink & Hare, 2016; Zhou et al., 2019).

As there is no formal goodness-of-fit test for robust design, data were collapsed into primary occasions and run as a fully parameterised Cormack–Jolly–Seber model to assess goodness of fit. This model was assessed using Program RELEASE Test 2 and Test 3, which test the assumptions that (a) the probability of recapture is the same across individuals and (b) individuals have the same probability of surviving between a given time interval, regardless of the occasion at which they were first caught (Burnham et al., 1987). Time intervals between primary occasions were formatted to output annual survival probabilities.

In total, 24 models were built with varying constraints on the real estimable parameters: probability of survival ϕ (which could be constant or time-varying); temporary emigration γ (which could follow a random or Markovian pattern or be absent); and the probability of capture c and recapture p (which could be equal to each other or distinct, constant or time-varying), Table S1. Corrected Akaike information criterion (AICc) values were calculated for each model and used as weightings for model selection. The analyses were run twice, once using the human-only data set and once using the AI-supplemented data set.

3 | RESULTS

3.1 | Image processing

Human-only image processing identified 402 unique individuals, whereas AI-supplemented image processing identified 360, Table 1. The verified data set inferred that our true sample consisted of 360 individuals. No false acceptances were detected in either the human-only or AI-supplemented method; however, 88 incidences of false rejections were highlighted within the human-only data set (FRR = 0.17, Table 1). These corresponded to 42 recaptured animals that had been assigned as new individuals, three of which were reassigned twice and one that was reassigned three times.

3.2 | Capture–recapture analyses

Both data sets met the assumptions of model fit (Table S2). Although AICc model rankings altered marginally between data sets, in both analyses, the model with the lowest AICc value included constant survival, equal time-varying probability of capture and recapture and random temporary emigration (Model 5, Table 2) and was selected

TABLE 1 The total number of discrete individuals identified by the human-only and AI-supplemented image processing methods, alongside their respective false acceptance rate (FAR) and false rejection rate (FRR).

	Human-only	AI-supplemented
Total individuals	402	360
FAR	0	0
FRR	0.17	0

BRITISH ECOLOGICAL SOCIETY Ecological Solutions and Evidence

as the best model for comparison. As delta AICc intervals were narrow for models run with the human-only data set, we further considered model-averaged estimates but found no significant changes to our results, Table S3.

The model estimated consistently lower population abundance values for each primary occasion when run with the AI-supplemented data set (Figure 2). Both data sets demonstrate an overall trend of increasing population size in the monitoring period following cyclone Fehi, with a notable increase after the first primary occasion.

TABLE 2 The top five models weighted by AICc for the human-only and AI-supplemented data sets.

Model	AICc	ΔAICc	AICc weights
Human-only			
5 $\phi_{(.)}\gamma'_{(t)} = \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	**4470.39**	**0**	**0.259**
3 $\phi_{(t)}\gamma' = \gamma'' = 0\ p_{(t)} = c_{(t)}$	4470.47	0.08	0.249
2 $\phi_{(t)}\gamma'_{(t)} = \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4470.65	0.26	0.228
4 $\phi_{(.)}\gamma'_{(t)} \neq \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4470.73	0.33	0.219
1 $\phi_{(t)}\gamma'_{(t)} \neq \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4473.90	3.50	0.045
AI-supplemented			
5 $\phi_{(.)}\gamma'_{(t)} = \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	**4511.12**	**0**	**0.742**
2 $\phi_{(t)}\gamma'_{(t)} = \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4514.27	3.15	0.154
4 $\phi_{(.)}\gamma'_{(t)} \neq \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4515.84	4.71	0.070
1 $\phi_{(t)}\gamma'_{(t)} \neq \gamma''_{(t)}\ p_{(t)} = c_{(t)}$	4517.77	6.65	0.0267
3 $\phi_{(t)}\gamma' = \gamma'' = 0\ p_{(t)} = c_{(t)}$	4520.29	9.17	0.008

Note: Survival ϕ, capture c and recapture p could be constant (.) or time-varying (*t*). $\gamma'_{(t)} = \gamma''_{(t)}$ denotes random movement, $\gamma'_{(t)} \neq \gamma''_{(t)}$ denotes Markovian movement and $\gamma' = \gamma'' = 0$ denotes no movement. The best model for both data sets (Model 5) is highlighted in bold.

TABLE 3 Estimated probability of temporary emigration γ, ±SE between each primary sampling occasion.

Period	Human-only	AI-supplemented
January 2019 to March 2019	0.03 ± 0.22	0 ± 0
March 2019 to January 2020	0.23 ± 0.16	0.21 ± 0.14
January 2020 to November 2020	0.48 ± 0.10	0.43 ± 0.10
November 2020 to February 2021	0.73 ± 0.07	0.63 ± 0.08

The annual survival probability predicted by the AI-supplemented data set (ϕ = 0.51 ± SE 0.07) was higher than predicted by the human-only data set (ϕ = 0.46 ± SE 0.08). The human-only data set predicted higher probabilities of temporary emigration between primary sampling occasions, Table 3. Capture probabilities across secondary occasions varied between 0.01 and 0.037 for each data set. A full table of capture probabilities can be found in the Supporting Information (Table S4).

4 | DISCUSSION

This study provides evidence that using AI software to identify individual animals can influence multiple parameter estimates in the demographic analysis of a wild population. Here, the use of the HotSpotter package specifically produced a data set with a lower FRR relative to human-only image processing, reducing the overinflation of population abundance estimates and underestimation of the survival rate. The results further indicate an overall pattern of population increase for our focal species, the Kapitia skink, and we suggest the use of the estimates produced by our AI-supplemented data set to inform its conservation.

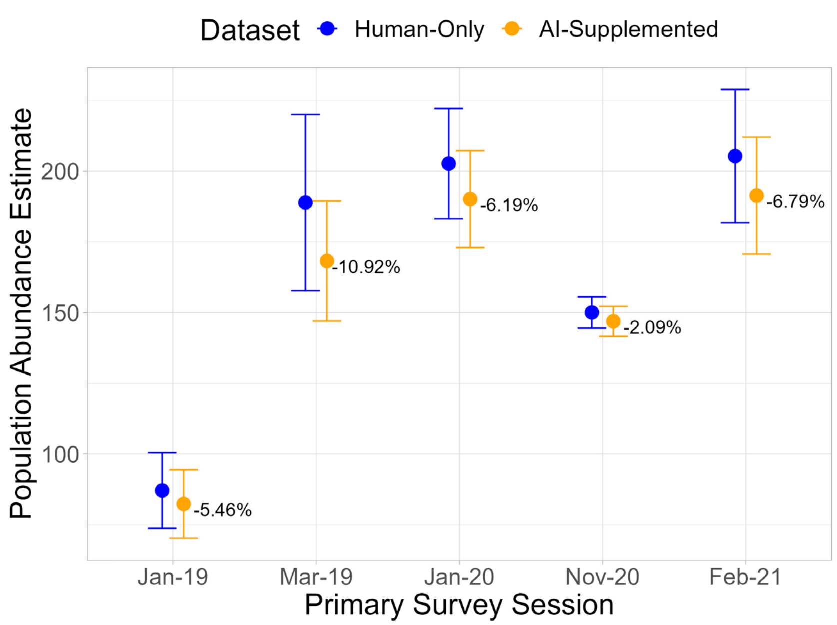

FIGURE 2 Comparison of population abundance estimates produced by analysis of the human-only data set (blue, left) and of the AI-supplemented data set (orange, right) with standard error bars. Percentage values show the difference in mean population estimates produced by the AI-supplemented data set relative to estimates produced by the human-only data set at each primary survey occasion.

Previous studies have concluded that multiple AI software packages can aid the recognition of individual animals (Bardier et al., 2020; Dalibard et al., 2021; Dunbar et al., 2021; Hou et al., 2020; Rotger et al., 2019) or have used known animal data or simulations to evidence that even small errors in encounter history data sets can bias demographic estimates (Bohnett et al., 2023; Johansson et al., 2020; Morrison et al., 2011; Rakhimberdiev et al., 2022; Yoshizaki et al., 2009). However, to our knowledge, this is the first in situ case study directly demonstrating the impact of using an AI image recognition software package on estimates that will influence an active conservation programme.

4.1 | Image processing

When comparing numerous images of individual animals purely 'by-eye', even experienced species practitioners are liable to error (Bohnett et al., 2023; Cruickshank & Schmidt, 2017). Here, the use of HotSpotter image recognition software highlighted several false rejections made by human observers, in which skinks were recorded as new individuals despite being captured on previous occasions. These missed matches likely occurred due to the difficult and fatiguing nature of comparing an image with a large catalogue of individual animals. In contrast, HotSpotter presented a narrowed field of probable matches to a target image, streamlining the process for observers to improve the probability that a recaptured individual would be identified.

No false acceptances, where unique individuals may be recorded as a single individual, were detected in either data set. False matches are more rarely reported in assessments of photographic individual identification approaches: They may be difficult to identify or a less common error (Bardier et al., 2020; Cruickshank & Schmidt, 2017; Johansson et al., 2020). It is plausible that human observers are more inclined to split individuals when it is felt a match is uncertain; discussions with observers suggested that wrongly matching two unique individuals could feel more like an 'active' rather than 'passive' mistake. In addition, research reliant on methods such as camera-trapping or sampling at a distance is likely to contend with inconsistent and poor quality images which distort an animal's recognisable features, making them difficult to attribute to previously caught individuals and therefore more likely to be falsely split (Kodi et al., 2024; Stevick et al., 2001).

It is important to acknowledge that this study could not rely upon a data set of known individuals to fully validate the accuracy of matches made; therefore, unidentified errors may still be present in both the human-only and AI-supplemented data sets. However, our verification process indicated that the AI-supplemented data set is a more accurate reflection of the true individual encounter histories: images assigned alternate identifications by the human-only and AI-supplemented approaches were further evaluated, and for each, the identification deemed accurate corresponded unanimously to the AI-supplemented data set. Furthermore, studies conducted using data sets of known individuals have consistently reported increased accuracy when AI software is used to assist individual identification (Bardier et al., 2020; Cruickshank & Schmidt, 2017; Dalibard et al., 2021; Dunbar et al., 2021; Hou et al., 2020; Rotger et al., 2019). We therefore advise that estimates produced by our AI-supplemented data set is favoured to inform population management of the Kapitia skink.

4.2 | Impact on parameter estimates

Parameter estimates differed dependent on the method used to identify individual capture histories. A FRR of 0.17 was observed in the human-only data set, greater than the 0.10 threshold suggested by Morrison et al. (2011) above which population estimates are likely to experience significant bias. In contrast, the AI-supplemented data set had an FRR of 0 and produced lower abundance estimates than the human-only data set, suggesting that relying on human observation alone may have resulted in overestimates of population size. As FRRs are higher in human observation methods than those that utilise AI, commonly >0.10, it is likely that human-only approaches are disposed to overestimating population abundance (Bardier et al., 2020; Cruickshank & Schmidt, 2017). Mechanistically, this divergence is a direct result of missed matches falsely inflating the number of observed individuals. Similarly, estimates of survival probability were higher when missed matches were accounted for in the AI-supplemented data set; the model otherwise considering that these individuals may have experienced mortality.

The biases demonstrated in our analysis correspond to those predicted by studies simulating errors in encounter histories and demonstrated in camera trap analyses of captive snow leopards (Bohnett et al., 2023; Johansson et al., 2020; Morrison et al., 2011; Renet et al., 2019; Yoshizaki, 2009). This study contributes new evidence that these effects are applicable to in situ conservation in divergent taxa.

These findings are significant as even small biases in survival and abundance estimates can have a substantial effect when managing threatened populations (Bickerton et al., 2023; Lloyd-Jones et al., 2023). Demographic information often informs planning interventions such as reintroductions, threat mitigation and habitat management, and skewed estimates may impact their effectiveness (Bird et al., 2021; Callaghan et al., 2024; Conde et al., 2019; Morrison et al., 2021; Volis & Deng, 2020). In addition, misevaluation of extinction risk can lead to poor resource allocation or a lack of urgency that could prove crucial to rare species. Underestimation of extinction risk is of particular concern as evidence suggests that 'splitting' errors, which inflate population estimates, are most common in capture–recapture data sets constructed from photographic identification (Bardier et al., 2020; Bohnett et al., 2023; Cruickshank & Schmidt, 2017; Johansson et al., 2020; Kodi et al., 2024). Due to the potential of AI to reduce bias in population estimates, it is desirable that more conservation projects consider implementing the use of image recognition software in monitoring programmes.

4.3 | Caveats and considerations

In this study, we used a single AI image recognition software, HotSpotter, in our AI-supplemented approach to identifying individual animals. It is possible that using alternate packages to process images of the Kapitia skink may have produced different outcomes, given that the underlying mechanics of image sorting can differ variously. However, given the overall trend towards FRR reduction seen across studies comparing relevant packages, it is reasonable to assume that the reduction of inflated estimates observed with HotSpotter would likely extend to comparable software.

Although software packages for individual identification are broadly efficient and accessible, projects should review the resources required for set-up and implementation. For instance, human input remains necessary to build a database of digital images, run the software and make final decisions on identification. Although improved automation of image data handling may become increasingly accessible to practitioners as machine learning frameworks continue to be developed (Bogucki et al., 2019; de Lorm et al., 2023; Hou et al., 2020; Norouzzadeh et al., 2018; Petso et al., 2022; Schneider et al., 2019).

Project managers should further consider which software is the most appropriate for use in the context of the project needs. Some initiatives develop and maintain efficient species-specific individual identification software with simplified interfaces, but these incur high costs that may be prohibitive to smaller programmes (McClure et al., 2020). Yet some free software can experience limited maintenance, and therefore reduced longevity, which limits its scope for longer term monitoring (Crall et al., 2013).

Finally, both human- and machine-induced errors can persist in AI-supplemented data sets: There is no guarantee that any field data, however processed, will be error free. Integrating a misidentification framework into capture–recapture models may improve the reliability of estimates in analyses where it is reasonable to assume that misidentification rates remain significant. Although conversely, when either recapture probabilities or misidentifications are small, simulations indicate that standard model approaches are more reliable (Morrison et al., 2011; Yoshizaki et al., 2009). In addition to our own, many studies assessing the accuracy of AI-supplemented individual identification have reported low rates of misidentification errors of <0.05, indicating that this approach could negate the need to incorporate misidentification into a model (Bardier et al., 2020; Cruickshank & Schmidt, 2017; Monnet et al., 2022; Nipko et al., 2020; Rotger et al., 2019). Reported accuracies, however, varied with factors such as the quality or consistency of images and the software used, and it is plausible that some taxa and survey methods could incur higher rates of misidentification; thus, the context of a data set should be considered when selecting the optimal approach for analysis.

5 | CONCLUSIONS

Our analyses provide a case study demonstrating that an AI tool can mitigate bias in demographic estimates for in situ conservation. Here, errors in a human-only individual identification data set led to the likely overestimation of population size for the critically endangered Kapitia skink, of up to ~11%, and we advise that urgent conservation intervention for this species is informed by the estimates produced using AI-supplemented individual identification. More generally, this study recommends that photographic individual identification methods for capture–recapture are best complemented with AI software to reduce false or missed matches that skew population estimates. Efforts should continue to facilitate the automation and accessibility of individual identification software to aid practitioners, with the potential to prevent the underestimation of extinction risk for threatened species.

AUTHOR CONTRIBUTIONS

Lynn K. Adams, Rod Hitchmough and Les R. Moran designed the survey methodology and collected the data. Emily A. Jordan conducted the data analysis and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

EJ was supported by the National Environment Research Council (award NE/S007229/1), and species monitoring was support by the Department of Conservation. We acknowledge Ngāti Waewae as kaitiaki for Kapitia skinks and as mana whenua for the takiwā Kapitia skinks inhabit. We would like to thank the Hokitika Office of the Department of Conservation for their support in all aspects of the work.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/2688-8319.70018.

DATA AVAILABILITY STATEMENT

Data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.bvq83bkkh (Jordan et al., 2025).

STATEMENT ON INCLUSION

Our study brings together authors from different countries, including scientists based in the country where the study was carried out.

RELEVANT GREY LITERATURE

You can find related grey literature on the following topics on Applied Ecology Resources: Population estimation, Reptile conservation, Robust design.

ORCID

Emily A. Jordan https://orcid.org/0009-0005-1713-7964
David J. Murrell https://orcid.org/0000-0002-4830-8966

REFERENCES

Bardier, C., Székely, D., Augusto-Alves, G., Matínez-Latorraca, N., Schmidt, B. R., & Cruickshank, S. S. (2020). Performance of visual vs. software-assisted photo-identification in mark-recapture studies: A case study examining different life stages of the pacific horned frog (*Ceratophrys stolzmanni*). *Amphibia-Reptilia*, *42*(1), 17–28. https://doi.org/10.1163/15685381-bja10025

Barker, R. J., & White, G. C. (2004). Towards the mother–of–all–models: Customised construction of the mark–recapture likelihood function. *Animal Biodiversity and Conservation*, *27*(1), 177–185.

Belaud, M., Bertolero, A., & Cheylan, M. (2022). Estimating the demographics of an ocellated lizard (*Timon lepidus* Daudin, 1802) population through photo identification capture-recapture. *Ecologia Mediterranea*, *48*(2), 23–32. https://doi.org/10.3406/ecmed.2022.2151

Bickerton, K. T., Ewen, J. G., Canessa, S., Cole, N. C., Frost, F., Mootoocurpen, R., & McCrea, R. (2023). Avoiding bias in estimates of population size for translocation management. *Ecological Applications*, *33*(8), e2918.

Bird, J. P., Woodworth, B. K., Fuller, R. A., & Shaw, J. D. (2021). Uncertainty in population estimates: A meta-analysis for petrels. *Ecological Solutions and Evidence*, *2*(3), e12077.

Bird, T., Lyon, J., Nicol, S., McCarthy, M., & Barker, R. (2014). Estimating population size in the presence of temporary migration using a joint analysis of telemetry and capture–recapture data. *Methods in Ecology and Evolution*, *5*(7), 615–625.

Black, I. R., Berman, J. M., Cadena, V., & Tattersall, G. J. (2019). Behavioral thermoregulation in lizards: Strategies for achieving preferred temperature. In J. M. Berman (Ed.), *Behavior of lizards* (pp. 13–46). CRC Press.

Bogucki, R., Cygan, M., Khan, C. B., Klimek, M., Milczek, J. K., & Mucha, M. (2019). Applying deep learning to right whale photo identification. *Conservation Biology*, *33*(3), 676–684.

Bohnett, E., Poya Faryabi, S., Lewison, R., An, L., Bian, X., Rajabi, A. M., Jahed, N., Rooyesh, H., Mills, E., Ramos, S., Mesnildrey, N., Santoro Perez, C. M., Taylor, J., Terentyev, V., & Ostrowski, S. (2023). Human expertise combined with artificial intelligence improves performance of snow leopard camera trap studies. *Global Ecology and Conservation*, *41*, e02350. https://doi.org/10.1016/J.GECCO.2022.E02350

Bolger, D. T., Morrison, T. A., Vance, B., Lee, D., & Farid, H. (2012). A computer-assisted system for photographic mark–recapture analysis. *Methods in Ecology and Evolution*, *3*(5), 813–822. https://doi.org/10.1111/j.2041-210X.2012.00212.x

Burnham, K. P., Anderson, D. R., White, G. C., Brownie, C., & Pollock, K. P. (1987). *Design and analysis of methods for fish survival experiments based on release–recapture*. Monograph 5. American Fisheries Society, Bethesda.

Callaghan, C. T., Santini, L., Spake, R., & Bowler, D. E. (2024). Population abundance estimates in conservation and biodiversity research. *Trends in Ecology & Evolution*, *39*, 515–523.

Conde, D. A., Staerk, J., Colchero, F., da Silva, R., Schöley, J., Maria Baden, H., Jouvet, L., Fa, J. E., Syed, H., Jongejans, E., Meiri, S., Gaillard, J. M., Chamberlain, S., Wilcken, J., Jones, O. R., Dahlgren, J. P., Steiner, U. K., Bland, L. M., Gomez-Mestre, I., … Vaupel, J. W. (2019). Data gaps and opportunities for comparative and conservation biology. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(19), 9658–9664. https://doi.org/10.1073/PNAS.1816367116/-/DCSUPPLEMENTAL

Crall, J. P., Stewart, C. v., Berger-Wolf, T. Y., Rubenstein, D. I., & Sundaresan, S. R. (2013). HotSpotter—Patterned species instance recognition. *2013 IEEE Workshop on Applications of Computer Vision (WACV)*. https://doi.org/10.1109/WACV.2013.6475023

Cruickshank, S. S., & Schmidt, B. R. (2017). Error rates and variation between observers are reduced with the use of photographic matching software for capture-recapture studies. *Amphibia-Reptilia*, *38*(3), 315–325. https://doi.org/10.1163/15685381-00003112

Dalibard, M., Buisson, L., Calvez, O., Nguyen-Hong, M., Trochet, A., & Laffaille, P. (2021). Can ventral pattern be used for individual recognition of the vulnerable Pyrenean brook newt (*Calotriton asper*)? *Herpetological Journal*, *43*, 99–110. https://doi.org/10.33256/31.2.99110

Dawson, J., Panter, C. T., & Zeisset, I. (2021). Comparisons of image-matching software when identifying pool frog (*Pelophylax lessonae*) individuals from a reintroduced population herpetologica. *Herpetological Journal*, *31*(1), 55–59. https://doi.org/10.33256/31.1.5559

de Lorm, T. A., Horswill, C., Rabaiotti, D., Ewers, R. M., Groom, R. J., Watermeyer, J., & Woodroffe, R. (2023). Optimizing the automated recognition of individual animals to support population monitoring. *Ecology and Evolution*, *13*(7), e10260.

Dunbar, S. G., Anger, E. C., Parham, J. R., Kingen, C., Wright, M. K., Hayes, C. T., Safi, S., Holmberg, J., Salinas, L., & Baumbach, D. S. (2021). HotSpotter: Using a computer-driven photo-id application to identify sea turtles. *Journal of Experimental Marine Biology and Ecology*, *535*, 151490. https://doi.org/10.1016/J.JEMBE.2020.151490

Fitzharris, B. (2007). How vulnerable is New Zealand to the impacts of climate change? *New Zealand Geographer*, *63*(3), 160–168. https://doi.org/10.1111/j.1745-7939.2007.00119.x

Hammond, P. S. (1990). Heterogeneity in the Gulf of Maine? Estimating humpback whale population size when capture probabilities are not equal. *Reports of the International Whaling Commission*, *12*, 135–139.

Hitchmough, R. (2021). *Oligosoma salmo*. The IUCN Red List of threatened species. https://www.iucnredlist.org/species/156730274/156730436

Hou, J., He, Y., Yang, H., Connor, T., Gao, J., Wang, Y., Zeng, Y., Zhang, J., Huang, J., Zheng, B., & Zhou, S. (2020). Identification of animal individuals using deep learning: A case study of giant panda. *Biological Conservation*, *242*, 108414. https://doi.org/10.1016/J.BIOCON.2020.108414

Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika*, *76*(1), 133–140. https://doi.org/10.1093/biomet/76.1.133

Huggins, R. M. (1991). Some practical aspects of a conditional likelihood approach to capture experiments. *Biometrics*, *47*(2), 725–732. https://doi.org/10.2307/2532158

Jain, A. (2007). Biometric recognition. *Nature*, *449*, 38–40. https://doi.org/10.1038/449038a

Johansson, Ö., Samelius, G., Wikberg, E., Chapron, G., Mishra, C., & Low, M. (2020). Identification errors in camera-trap studies result in systematic population overestimation. *Scientific Reports*, *10*(1), 6393.

Jordan, E. A., Ewen, J., Hitchmough, R., Moran, L. R., Murrell, D. J., Rowcliffe, M. J., & Adams, L. K. (2025). Data from: Supplementing human observation with artificial intelligence impacts demographic estimates for a Critically Endangered lizard. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.bvq83bkkh

Kodi, A. R., Howard, J., Borchers, D. L., Worthington, H., Alexander, J. S., Lkhagvajav, P., Bayandonoi, G., Ochirjav, M., Erdenebaatar, S., Byambasuren, C., Battulga, N., Johansson, Ö., & Sharma, K. (2024). Ghostbusting—Reducing bias due to identification errors in spatial capture-recapture histories. *Methods in Ecology and Evolution*, *15*(6), 1060–1070. https://doi.org/10.1111/2041-210X.14326

Lettink, M., & Hare, K. M. (2016). Sampling techniques for New Zealand lizards. In D. Chapple (Ed.), *New Zealand lizards* (pp. 269–291). Springer. https://doi.org/10.1007/978-3-319-41674-8_10/COVER

Lloyd-Jones, L. R., Bravington, M. V., Armstrong, K. N., Lawrence, E., Feutry, P., Todd, C. M., & Westcott, D. A. (2023). Close-kin mark-recapture informs critically endangered terrestrial mammal status. *Scientific Reports*, *13*(1), 12512.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

McCann, S., & Lowe, D. G. (2012). Local naive Bayes nearest neighbor for image classification. In *2012 IEEE conference on computer vision and pattern recognition* (pp. 3650–3656). IEEE. https://doi.org/10.1109/CVPR.2012.6248111

McClure, E. C., Sievers, M., Brown, C. J., Buelow, C. A., Ditria, E. M., Hayes, M. A., Pearson, R. M., Tulloch, V. J. D., Unsworth, R. K. F., & Connolly, R. M. (2020). Artificial intelligence meets citizen science to supercharge ecological monitoring. *Patterns*, *1*(7), 100109. https://doi.org/10.1016/j.patter.2020.100109

Melzer, S., Hitchmough, R. A., Bell, T., Chapple, D. G., & Patterson, G. B. (2019). Lost and found: Taxonomic revision of the speckled skink (*Oligosoma infrapunctatum*; Reptilia; Scincidae) species complex from New Zealand reveals a potential cryptic extinction, resurrection of two species, and description of three new species. *Zootaxa*, *4623*(3), 441–484. https://doi.org/10.11646/zootaxa.4623.3.2

Monnet, C., Dokhelar, T., & Renet, J. (2022). Rapid colour changes in a tiny threatened gecko do not impede computer-assisted individual recognition. *bioRxiv*, https://doi.org/10.1101/2022.03.16.484634

Morrison, C. A., Butler, S. J., Robinson, R. A., Clark, J. A., Arizaga, J., Aunins, A., & Gill, J. A. (2021). Covariation in population trends and demography reveals targets for conservation action. *Proceedings of the Royal Society B*, *288*(1946), 20202955.

Morrison, T. A., Yoshizaki, J., Nichols, J. D., & Bolger, D. T. (2011). Estimating survival in photographic capture-recapture studies: Overcoming misidentification error. *Methods in Ecology and Evolution*, *2*(5), 454–463. https://doi.org/10.1111/j.2041-210X.2011.00106.x

Moya, O., Mansilla, P. L., Madrazo, S., Igual, J. M., Rotger, A., Romano, A., & Tavecchia, G. (2015). APHIS: A new software for photo-matching in ecological studies. *Ecological Informatics*, *27*, 64–70.

Nelson, N. J., Romijn, R. L., Dumont, T., Reardon, J. T., Monks, J. M., Hitchmough, R. A., Empson, R., & Briskie, J. V. (2016). Lizard conservation in mainland sanctuaries. In *New Zealand lizards* (pp. 321–339). Springer International Publishing. https://doi.org/10.1007/978-3-319-41674-8_12

Nipko, R. B., Holcombe, B. E., & Kelly, M. J. (2020). Identifying individual jaguars and ocelots via pattern-recognition software: Comparing HotSpotter and wild-ID. *Wildlife Society Bulletin*, *44*(2), 424–433. https://doi.org/10.1002/WSB.1086

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716–E5725. https://doi.org/10.1073/PNAS.1719367115

Park, H., Lim, A., Choi, T. Y., Baek, S. Y., Song, E. G., & Park, Y. C. (2019). Where to spot: Individual identification of leopard cats (*Prionailurus bengalensis euptilurus*) in South Korea. *Journal of Ecology and Environment*, *43*(1), 1–5. https://doi.org/10.1186/S41610-019-0138-Z/FIGURES/3

Petso, T., Jamisola, R. S., & Mpoeleng, D. (2022). Review on methods used for wildlife species and individual identification. *European Journal of Wildlife Research*, *68*(1), 1–18. https://doi.org/10.1007/S10344-021-01549-4 /TABLES/3.

Rakhimberdiev, E., Karagicheva, J., Saveliev, A., Loonstra, A. H. J., Verhoeven, M. A., Hooijmeijer, J. C. E. W., Schaub, M., & Piersma, T. (2022). Misidentification errors in reencounters result in biased estimates of survival probability from CJS models: Evidence and a solution using the robust design. *Methods in Ecology and Evolution*, *13*(5), 1106–1118. https://doi.org/10.1111/2041-210X.13825

Renet, J., Leprêtre, L., Champagnon, J., & Lambret, P. (2019). Monitoring amphibian species with complex chromatophore patterns: A non-invasive approach with an evaluation of software effectiveness and reliability. *Herpetological Journal*, *29*(1), 13–22. https://doi.org/10.33256/hj29.1.1322

Rotger, A., Colomar, V., Moreno, J. E., & Parpal, L. (2019). Photo-identification of horseshoe whip snakes (*Hemorrhois hippocrepis*, Linnaeus, 1758) by a semi-automatic procedure applied to wildlife management. *Herpetological Journal*, *29*(4), 304–307. https://doi.org/10.33256/hj29.4.304307

Safaei Pirooz, A. A., Flay, R. G., Turner, R., & Azorin-Molina, C. (2019). Effects of climate change on New Zealand design wind speeds. *National Emergency Response*, *32*(2), 14–20.

Schneider, S., Taylor, G. W., Linquist, S., & Kremer, S. C. (2019). Past, present and future approaches using computer vision for animal re-identification from camera trap data. *Methods in Ecology and Evolution*, *10*(4), 461–470. https://doi.org/10.1111/2041-210X.13133

Stevick, P. T., Palsbøll, P. J., Smith, T. D., Bravington, M. V., & Hammond, P. S. (2001). Errors in identification using natural markings: Rates, sources, and effects on capture recapture estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, *58*(9), 1861–1870. https://doi.org/10.1139/f01-131

Strampelli, P., Searle, C. E., Smit, J. B., Henschel, P., Mkuburo, L., Ikanda, D., Macdonald, D. W., & Dickman, A. J. (2022). Camera trapping and spatially explicit capture–recapture for the monitoring and conservation management of lions: Insights from a globally important population in Tanzania. *Ecological Solutions and Evidence*, *3*(1), e12129. https://doi.org/10.1002/2688-8319.12129

Tucker, A. M., McGowan, C. P., Robinson, R. A., Clark, J. A., Lyons, J. E., Derose-Wilson, A., du Feu, R., Austin, G. E., Atkinson, P. W., & Clark, N. A. (2019). Effects of individual misidentification on estimates of survival in long-term mark–resight studies. *The Condor*, *121*(1), duy017. https://doi.org/10.1093/CONDOR/DUY017

van Tienhoven, A. M., den Hartog, J. E., Reijns, R. A., & Peddemors, V. M. (2007). A computer-aided program for pattern-matching of natural marks on the spotted Raggedtooth shark *Carcharias tuaurus*. *Journal of Applied Ecology*, *44*, 273–280.

van Winkel, D., Baling, M., & Hitchmough, R. (2018). *Reptiles and amphibians of New Zealand: A field guide*. Bloomsbury.

Vidal, M., Wolf, N., Rosenberg, B., Harris, B. P., & Mathis, A. (2021). Perspectives on individual animal identification from biology and computer vision. *Integrative and Comparative Biology*, *61*(3), 900–916. https://doi.org/10.1093/ICB/ICAB107

Volis, S., & Deng, T. (2020). Importance of a single population demographic census as a first step of threatened species conservation planning. *Biodiversity and Conservation*, *29*(2), 527–543.

White, G. C., & Burnham, K. P. (1999). Program MARK: Survival estimation from populations of marked animals. *Bird Study*, *46*(Supplement), 120–138.

Yoshizaki, J., Pollock, K. H., Brownie, C., & Webster, R. A. (2009). Modeling misidentification errors in capture–recapture studies using photographic identification of evolving marks. *Ecology*, *90*(1), 3–9. https://doi.org/10.1890/08-0304.1

Zhou, L., Liang, T., & Shi, L. (2019). Amphibian and reptilian chorotypes in the arid land of Central Asia and their determinants. *Scientific Reports*, *9*(1), 9453.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Table S1: List of candidate models. Survival ϕ, capture p, and recapture c, could be constant (.) or vary with time (t).

Table S2: Results of Program RELEASE goodness of fit tests 2, 3, and combined 2+3 for each data set analysed.

Table S3: A comparison of model-averaged and model 5 results produced by the Human-Only data set.

Table S4: The capture probabilities for each secondary occasion across all primary occasions, produced by the analysis of the 'Human-Only' and 'AI-Supplemented' data sets.

How to cite this article: Jordan, E. A., Ewen, J., Hitchmough, R., Moran, L. R., Murrell, D. J., Rowcliffe, J. M., & Adams, L. K. (2025). Supplementing human observation with artificial intelligence impacts demographic estimates for a critically endangered lizard. *Ecological Solutions and Evidence*, *6*, e70018. https://doi.org/10.1002/2688-8319.70018