Herpetological Journal

Monitoring amphibian species with complex chromatophore patterns: a non-invasive approach with an evaluation of software effectiveness and reliability

Julien Renet[1], Lisa Leprêtre[1], Jocelyn Champagnon[2] & Philippe Lambret[2]

[1]Conservatoire d'espaces naturels de Provence-Alpes-Côte d'Azur, Pôle Biodiversité régionale, appartement n°5, 96 rue Droite, F-04200 Sisteron, France

[2]Tour du Valat, Research Institute for the Conservation of Mediterranean Wetlands, Le Sambuc, F-13200 Arles, France

The estimation of demographic parameters in wild populations is strengthened by individual identification. For amphibians, various techniques are used to either temporarily or permanently mark individuals for identification. Photo-identification of body patterns offers a non-invasive technique. However, the reliability of photo-recognition software is key to the reliable estimation of the true demographic parameters. In the current study, we assessed the effectiveness of fully-automated and semi-automated software: Wild-ID and APHIS. We used the cryptic salamander *Hydromantes strinatii* as our study species. We used the False Rejection Rate (FRR) of Top 1, Top 5 and Top 10 matches of chest and cloaca pictures. Finally, we assessed the bias induced by our FRR for the estimation of population size through simulation. Wild-ID FRRs ranged from 0.042 to 0.093 while APHIS' ranged from 0.227 to 0.547. Wild-ID was equally efficient with pictures from the chest and from the cloaca, while APHIS was significantly more efficient with chest pictures than cloaca pictures. Cropping pictures did not significantly improve Wild-ID effectiveness. Our Wild-ID FRRs are among the lowest ever obtained from pictures of an amphibian with a complex chromatophore pattern. Simulation showed that the Top 10 FRR from selected software Wild-ID induced a low bias 2.7% on the estimation of population size. The effectiveness and plasticity of Wild-ID provides opportunities for reliably monitoring amphibian species with complex colour patterns.

Key words: Photographic identification, Wild-ID, APHIS, False Rejection Rate, *Hydromantes strinatii*, demographic parameters, monitoring

INTRODUCTION

The assessment of demographic parameters using standardised methods is a key part of conservation biology, especially to understand population trends and make reliable predictions about their viability (Griffiths et al., 2015). Capture – mark – recapture models (CMR) that estimate population size, survival and detectability are useful tools for obtaining data for conservation (McCrea & Morgan, 2014). This method requires that individual marks are perennial in order to allow individual recognition over time (Chao, 1989; Nichols, 1992). For amphibians, marking techniques can be either invasive or non-invasive. For some species, evident and consistent colour patterns such as spots or stripes can be used as natural marks for individual recognition (Arntzen et al., 2004; Wengert & Gabrial, 2006; Ferner, 2010). Although visual matching can be an efficient method for monitoring a small population over a limited time, this technique becomes tedious and time-consuming (and thus sensitive to mistakes) when used to monitor large

populations over several years (Morrison et al., 2011; Cruickshank & Schmidt, 2017).

The development of algorithms for pattern recognition allows pattern mapping, and therefore pattern matching, to be automated (Sacchi et al., 2016). However, two types of error are likely to induce a bias in the assessment of demographic parameters (Yoshizaki et al., 2009; Morrison et al., 2011). The first, namely False Acceptance (FA), consists of assigning the same identification to two different individuals. The second, False Rejection (FR), consists of the failure of software to match two pictures of the same individual (Jain, 2007). The rates at which these two types of error occur (respectively FAR and FRR) can be used to assess software effectiveness, the latter being more commonly used (Bolger et al., 2012).

There are three types of software that can be distinguished: (1) software, such as I3S series (Van Tienhoven et al., 2007) and AmphIdent (Matthé et al., 2008), that require laborious and time consuming pre-processing treatment, which consists of tracing (i.e. with the computer's mouse) the pattern outlines on each

Correspondence: Julien Renet (julien.renet@cen-paca.org)

picture, (2) software not requiring any pre-processing of pictures (except the recommended cropping of pictures), and consequently time-saving, such as Wild-ID (Bolger et al., 2012) and Hotspotter (Crall et al., 2013), and (3) intermediate software, such as APHIS (Moya et al., 2015), which has been recently developed especially for herpetofauna, i.e., with a reduced pre-processing approach consisting of identifying two landmarks on each picture that the software can use as reference points (e.g. the base of left and right hind legs). Type 1 software is only used for species with gross body patterns (e.g. *Triturus carnifex*, *Paramesotriton hongkongensis*) because patterns need to be characterised easily (Fu et al., 2013; Sannolo et al., 2016). Type 2 requires a good quality of individual pictures, otherwise the accuracy of recognition may not be reliable (Morrison et al., 2016). While there is most probably a continuum rather than a true dichotomy in the specific pattern complexity, the user-friendly software Wild-ID is generally used for the identification of amphibians showing large and/or contrasted spots (e.g. Caorsi et al., 2012; Elgue et al., 2014; Morrison et al., 2016; Romiti et al., 2016), but has been assessed only once with a complex colour (i.e. fine spotted) pattern, using a neotenic salamander (Bendik et al., 2013). Also, the lowest FRRs were obtained using pictures taken with a "digital single-lens reflex" camera (DSLR) and in highly standardised conditions (e.g. cropped pictures, laboratory conditions) (Mettouris et al., 2016). Since APHIS benefits from two reference spots indicated by hand, it theoretically allows for a lower standardisation level compared to fully automated software and could be a useful compromise between type 1 and 2 software in terms of time and handling needed to process the pictures. Yet to date, APHIS has rarely been used for amphibian monitoring (but see Romano et al., 2017, 2018 with *Salamandrina perspicillata*).

In this study, we used Wild-ID and APHIS to identify the French cave salamanders *Hydromantes strinatii*. This species bears a complex chromatophore network resulting in a stippled colour pattern and thus provides a useful study species for testing software effectiveness and reliability. To assess software effectiveness (i.e. the FRR values) and the reliability of the entire photographic identification process (i.e. the FRR variability), we took pictures in field conditions with a "point and shoot" camera; in other words, we rejected the option of using a heavy-handling set-up in the field. We considered two different regions of the salamander body (the chest and the cloaca) to identify differences in FRR between body sites. We first compared the respective FRR and, secondly, we assessed the effect of the lowest FRR on the estimation of population size, as demographic parameters are the ultimate goals of CMR. This allowed us to determine whether the best combination of software and body region lead to acceptable errors in this estimation.

MATERIALS AND METHODS

Study species
Hydromantes strinatii is a salamandrid endemic to south-eastern France (Alpes-Maritimes and Alpes-de-Haute-Provence departments) and north-western Italy (Ligure region). It has a complex chromatophore pattern with a discrete phenotypic variability in its design and thus, it is a good model for the assessment of non-invasive identification methods, such as photo-identification, used to compute the demographic parameters needed for population viability analyses (Fig. 1). Further, it is important to study this species because it has been evaluated as "Near Threatened" and is therefore of conservation interest (Temple & Cox, 2009; Renet & Delauge, 2012). Its snout-vent length does not exceed ca. 75-80 mm, while maximal total length is 123 mm (Lanza et al., 1995). Individuals shy away from daylight in natural and artificial cavities. During the night, they come out and can be found in epigeous rocky habitats.

Population sampling and individual marking
During April 2015, we sampled a population in an epigeic habitat south of Roquebillière city (43°59'50.05''N, 7°18'41.57''E, Alpes-Maritimes department). We searched for individuals on six occasions with an equal sampling effort along a 94 m section of retaining walls, during the maximal activity period of the species, i.e., during nights with relative humidity ranging from 70 % to 100 % and temperature ranging from 5 °C to 15 °C (Lanza et al., 2006). As we repeatedly sampled the same section of wall, some individuals were captured on more than one sampling occasion (i.e. recaptured). Pattern resolution (i.e. scale) is likely to vary with body size and hence with individual growth. In order to avoid this potential bias, we captured adults only and avoided juveniles. Each individual was stored in a 5×10×10 cm box, with holes so that it could breathe, until it could be photographed. At the end of each sampling session, the animals were moved to the nearby photography setup. This consisted of a box lined on the bottom with polypropylene foam and equipped with a swinging glass plate. The animals were placed one by one upside down on the foam and held in place with the glass plate so that their ventral surface was flattened against the glass. For each individual, we took 4-5 pictures of the chest and 4-5 of the cloaca (Fig. 1). They were released immediately after photography at the exact location of capture, which was marked on each storage box. Manipulation of each individual did not exceed 5 minutes and total storage duration did not exceed 2 hours. Permission for this programme was issued by order of the Prefet [2015-227], according to French law.

Pictures were taken with a TG-3 Olympus© digital camera "point and shoot" (sensor: BSI CMOS 16 Mpx; zoom: 4 × 25-100mm f/2-4.9). We used the camera under the automatic "super-macro" mode with an additional LED LG-1 ring, which allowed the pictures to be taken 1 cm from the individuals. Pictures were taken with a 3456 × 4608 px resolution. No further processing was done, not even cropping; however, we also assessed whether cropping improved Wild-ID effectiveness (see below).

For each combination of sampling occasion x individual × body region, we selected the best picture among the 4-5 we had taken, thereby obtaining n_1 = 314 pictures for each
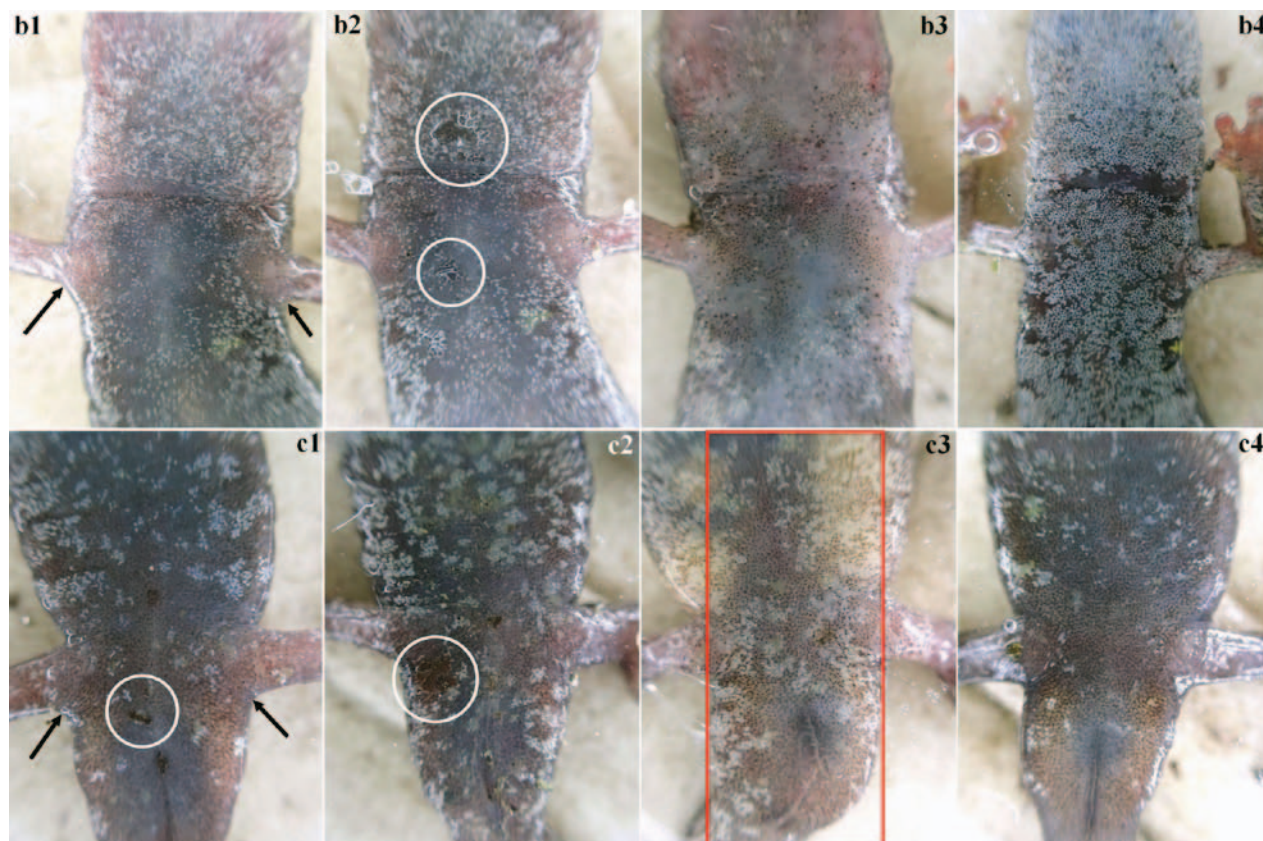
Figure 1. Variability of the complex chromatophore pattern of the French cave salamander *Hydromantes strinatii*, photographed at the two body regions considered in this study: chest (b) and cloaca (c). Picture b1 and b2 concern an individual during capture and recapture, respectively, the latter being identified as the first rank matching picture by Wild-ID during the five replicates despite minimal-standardisation conditions (see impurities and air bubbles indicated by circles). Black arrows indicate the armpit (b1) and the groin (c1) which we used as reference spots required by APHIS ITM approach. Red lines indicate the frame (1700 × 4608 px) used during the cropping test.

body region. We then allocated each pair of chest and cloaca pictures, among which some were capture pictures and others were recapture pictures, to an individual animal by visual matching, which we assume to be 100 % effective in discriminating each individual (i.e. FRR = 0). Indeed, the 314 pairs of chest and cloaca pictures were examined by two naive and independent observers who both recognised the same 253 individuals. There were 314 − 253 = 61 recaptures of 52 individuals; 61 − 52 = 9 and 52 − (61 − 52) = 43 individuals were, respectively, recaptured twice and once, and therefore photographed at three and two occasions; 314 − (43 × 2 + 9 × 3) = 201 individuals were captured and photographed once. We accordingly attributed an identification number (Id) to each of the 52 recaptured individuals.

Software characteristics and effectiveness assessment
APHIS (ver. 1.0; Moya et al., 2015) is available for free from https://imedea.uib-csic.es/bc/gep/docs/aphis/APHISPROGRAM/program/. It offers two individual matching processes: the Spots Pattern Matching (SPM) and the Image Template Matching (ITM). The former is based on I3S algorithm (Van Tienhoven et al., 2007) while the latter is based on algorithms developed by OpenCV (Open source Computer Vision). In this study, we used the ITM process, which requires the definition of two landmarks on the body of each individual; we used the

armpit and the vent for the chest and the cloaca regions respectively (Fig. 1: arrows). From there, the software creates a rectangle divided into six areas, which are analysed independently (Moya et al., 2015).

Wild-ID (ver. 1.0.1; Bolger et al., 2012) is available for free from http://software.dartmouth.edu/Macintosh/Academic/Wild-ID_1.0.0.zip. It uses the SIFT algorithm (Scale Invariant Feature Transform; Lowe, 2004) to extract the body pattern, and then compares the geometric arrangement of the SIFT features for each couple of pictures (i.e. matches) without pre-processing.

APHIS and Wild-ID offer, respectively, 100 and 20 "top-ranked" matches following a decreasing calculated score provided by their algorithms (i.e. the picture proposed at the 1st rank is that from the dataset showing the highest score with the submitted picture). One has to then search visually for the correct individual among the software proposals.

We assessed effectiveness of each software for each body region. For this, we determined, for each of the 61 recaptures, the rank at which the software proposed the correct picture (i.e. the 100 % true matched picture, identified from the visual assessment) among the top-ranked matches. Hence, we were able to state whether the software programme failed to rank each of the 61 recapture pictures within the Top 1, 5 or 10 matches; in other words, failed to correctly identify the

15

recaptured individuals in the first one, five or ten hits. Corresponding error rates were reported in terms of FRR (Jain, 2007). Both APHIS and Wild-ID compare one submitted picture to all pictures recorded in the dataset. During the matching process for APHIS and Wild-ID, each new recapture picture is compared against all pictures that have been processed previously; this means, for instance, that the first recapture pictures we processed was compared to the 253 pictures of 1st capture, and that the 61st recapture picture we processed has been compared to the 253 + 60 = 313 pictures of both 1st captures and recaptures. This means that the size of the dataset each new recapture picture is compared to increases for each subsequent picture. When assessing the FRR according to this method, i.e. by comparing one picture only to other pictures that had been processed so far rather than against all other existing pictures (all capture and recapture picture), the false rejection assignation, and therefore the FRR, may be strongly affected by the order of processing of the pictures. To take this potential bias into account, we decided to assess the software reliability, i.e. the variability of the FRR by repeating the process five times (i.e. five replicates). For each replicate the recapture pictures were processed in a different random order. Each of the three Top rank datasets therefore contained 1220 data points (5 replicates × 2 softwares × 2 body regions × 61 recapture pictures).

Wild-ID manual recommends to crop pictures so as to remove the background as much as possible, however, this can be a time consuming and laborious process. Thus, we assessed whether cropping was require to improve effectiveness. For this, we cropped cloaca pictures using a 1700 × 4608 px frame (Fig.1). We then randomly selected 30 recapture pictures and compared each of these to a database composed by the 253 + 60 = 313 pictures of both 1st captures and recaptures; for each comparison, the database was therefore renewed and slightly different from the previous (i.e. by one picture out of 313). This was done with the cropped pictures and with the original pictures, using the same 30 randomly selected individuals.

Statistical analysis

Failures, i.e. false rejections, were coded as 1 and successes coded as 0 in a new data frame, with software, body region and replicate as co-variables. We used generalised linear mixed models to test differences of FRR (dependent variable) between software and between body regions (explanatory variables, i.e. fixed factors). Mixed models were used to account for (1) the pseudo-replication induced by the fact that some individuals were recaptured and photographed more than once (Hurlbert, 1984) and (2) the replication scheme of the FRR computing. "Id" and "Replicate" were thus included as random factors. To assess whether FRR was influenced by the software and/or body region, we used each Top rank match and analysed the datasets using a binomial error using the following function:

$$logit(FRR) = \alpha + \beta_1.Soft + \beta_2.Reg + \beta_3.(Soft * Reg) + \varepsilon_{Id} + \varepsilon_{Repl}$$

where Soft is the software, Reg is the body region and Soft*Reg is their interaction, Id is this individual and Repl is the replicate. Contrast analyses were performed in order to assess whether the differences between the best combination of software and body region was significantly different from other combinations or not. To assess the variability of the FRR regarding the order in which recapture pictures were processed in, we computed the FRR for each replicate independently using the following function:

$$logit(FRR_i) = \alpha_i + \beta_1^i.Soft + \beta_2^i.Reg + \beta_3^i.(Soft * Reg) + \varepsilon_{Id}^i$$

where i is the replicate number. We then computed the standard deviation (SD) of the 5 FRR we obtained. Results are given as FRR estimated by the models together with their 95 % CI. Significance of fixed factors is given by a range of corresponding estimates ± standard error which does not overlap zero (i.e. meaning a null effect of the fixed factor). All analyses described above were performed in R 3.4.4 (R Core Team, 2018) using packages 'lme4' (Bates et al., 2014), 'lmerTest' (Kuznetsova et al., 2017) and 'effects' (Fox et al., 2016).

To assess the bias on the estimation of population size induced by the best software and body region FRRs obtained for each of the Top 1 and 10 matches, 1000 CMR datasets were simulated to which we applied the FRRs obtained from the generalised linear mixed model. The datasets were created assuming a population of 600 individuals, six occasions of sampling with a mean capture probability of 0.1 varying randomly for each sampling occasion. Those assumptions were based on the results of Mt standard closed capture model (Otis et al., 1978) applied to the real dataset with a time effect on detection probability and run in a Bayesian framework (population size = 543 ± 44 SD; mean detection probability = 0.098 ± 0.050 SD). The same Mt closed capture model was then applied to the 1000 simulated datasets with respectively no FRR, Top10 and Top1 FRRs. MCMC was applied using JAGS (Plummer, 2003) via the package R2jags (Su & Yajima, 2012) in R (R Core Team, 2018). We used a burn-in of 500 iterations, three chains, a thinning rate of two and 2000 iterations for each posterior distribution. Convergence was reached and mixing of the chains was good.

RESULTS

Fixed effects driving False Rejection Rate

For all three types of Top rank matches ($n_1 = n_2 = n_3 = 1220$), the FRR was significantly influenced by software (Top 1: |estimate| = 1.640 ± 0.234; Top 5: |estimate| = 1.909 ± 0.273; Top 10: |estimate| = 1.891 ± 0.286) and body region (Top 1: |estimate| = 0.826 ± 0.192; Top 5: |estimate| = 0.868 ± 0.196; Top 10: |estimate| = 0.914 ± 0.200) and their interaction (Top 1: |estimate| = 0.961 ± 0.324; Top 5: |estimate| = 0.596 ± 0.360; Top 10: |estimate| = 0.914 ± 0.390). Hence, we hereafter provide the FRR of each combination of software and body region separately.

Table 1. Software effectiveness: False Rejection Rate (FRR) when matching individual pictures of the French cave salamander *Hydromantes strinatii*, comparing software programmes and body regions. Contrast analyses were made between (1) the smallest FRR and others, and (2) between APHIS FRRs: ns = not significant (p > 0.05), * = significant (p < 0.05), *** = highly significant (p < 0.001).

High-scoring matches	Software	Body region	FRR estimated	Contrast analysis (1)	Contrast analysis (2)
Top 1	Wild-ID	Cloaca	0.082 (0.055-0.122)		
		Chest	0.093 (0.063-0.135)	|z| = 0.519 p = 0.604 ns	
	APHIS	Cloaca	0.547 (0.472-0.620)	|z| = 10.329 p < 0.001 ***	
		Chest	0.346 (0.279-0.419)	|z| = 7.363 p < 0.001 ***	|z| = 4.298 p < 0.001 ***
Top 5	Wild-ID	Chest	0.050 (0.030-0.081)		
		Cloaca	0.064 (0.041-0.100)	|z| = 1.008 p = 0.313 ns	
	APHIS	Chest	0.261 (0.203-0.328)	|z| = 7.702 p < 0.001 ***	
		Cloaca	0.457 (0.385-0.531)	|z| = 10.922 p < 0.001 ***	|z| = 5.058 p < 0.001 ***
Top 10	Wild-ID	Cloaca	0.042 (0.025-0.071)		
		Chest	0.042 (0.025-0.071)	|z| = 0 p = 1 ns	
	APHIS	Cloaca	0.422 (0.351-0.396)	|z| = 9.529 p < 0.001 ***	
		Chest	0.227 (0.173-0.291)	|z| = 6.613 p < 0.001 ***	|z| = 5.244 p < 0.001 ***

False Rejection Rate and its variability: the effects of software and body region

Wild-ID produced the smaller FRR for Top 1, Top 5 and Top 10 matches i.e. 0.082, 0.050 and 0.042, respectively (Table 1). These FRR values were significantly smaller than those of APHIS, for both chest and cloaca. Wild-ID FRR values were not significantly different for the two body regions (Table 1). In contrast, APHIS FRRs were significantly smaller for the chest than the cloaca region for Top 1, Top 5 and Top 10 matches. The standard deviation of the FRR across the five replicates ranged from 0.004 to 0.057 (Fig. 2). The FRR variability regarding the order of processing of the recapture pictures (i.e. across replicates) was greater with APHIS (SD range: 0.022–0.057) than with Wild-ID (SD range: 0.004–0.016).

Influence of cropping of pictures on Wild-ID effectiveness

Cropping did not significantly affect Wild-ID effectiveness: removing the background reduced the rank at which Wild-ID proposed the correct picture in only two recapture pictures out of 30: 7th instead of 8th rank and 18th instead of >20th rank. Hence, when coding these data as binary (false rejection or success), the two new data frames for cropped and original pictures were absolutely similar for the Top 1, Top 5 and Top 10 matches.

Influence of error in Wild-ID on the estimation of population size

With the selected software programme (Wild-ID) used on chest, FRR had a weak influence on the estimation of population size of *Hydromantes strinatii* (Table 3). Mean estimated population size for Control, Top 10 and Top 1 matches (the latter being the highest FRR and therefore inducing the strongest bias during the estimation of demographic parameters) all included the known population size of 600 individuals within their 95% confidence interval. Mean population size estimated from simulated data without FRR was underestimated by 2.0% [-16.6;14.5] while mean estimated population sizes were overestimated by 2.7% [-12.9;18.7] using Top 10 FRR, and by 9.0% [-7.7;25.8] using the Top 1 FRR.

DISCUSSION

Software effectiveness in recapture identification

Our results, obtained in field conditions, showed that Wild-ID was more effective and reliable than APHIS in matching pictures of a complex chromatophore pattern salamander (Table 1). Despite the field conditions and the use of a point and shoot camera, our Top 10 Wild-ID FRR values are among the lowest ever obtained and our more stringent Top 1 Wild-ID FRR values are close to the first quartile (i.e. among the lowest) of those found throughout the literature (Table 2). These low values are undoubtedly influenced by the small size of our database (*N* = 253 captures + 61 recaptures = 314) but there is probably also some influence of technological progress in the quality of sensor and optic lenses since the other studies were conducted, which helps to cope
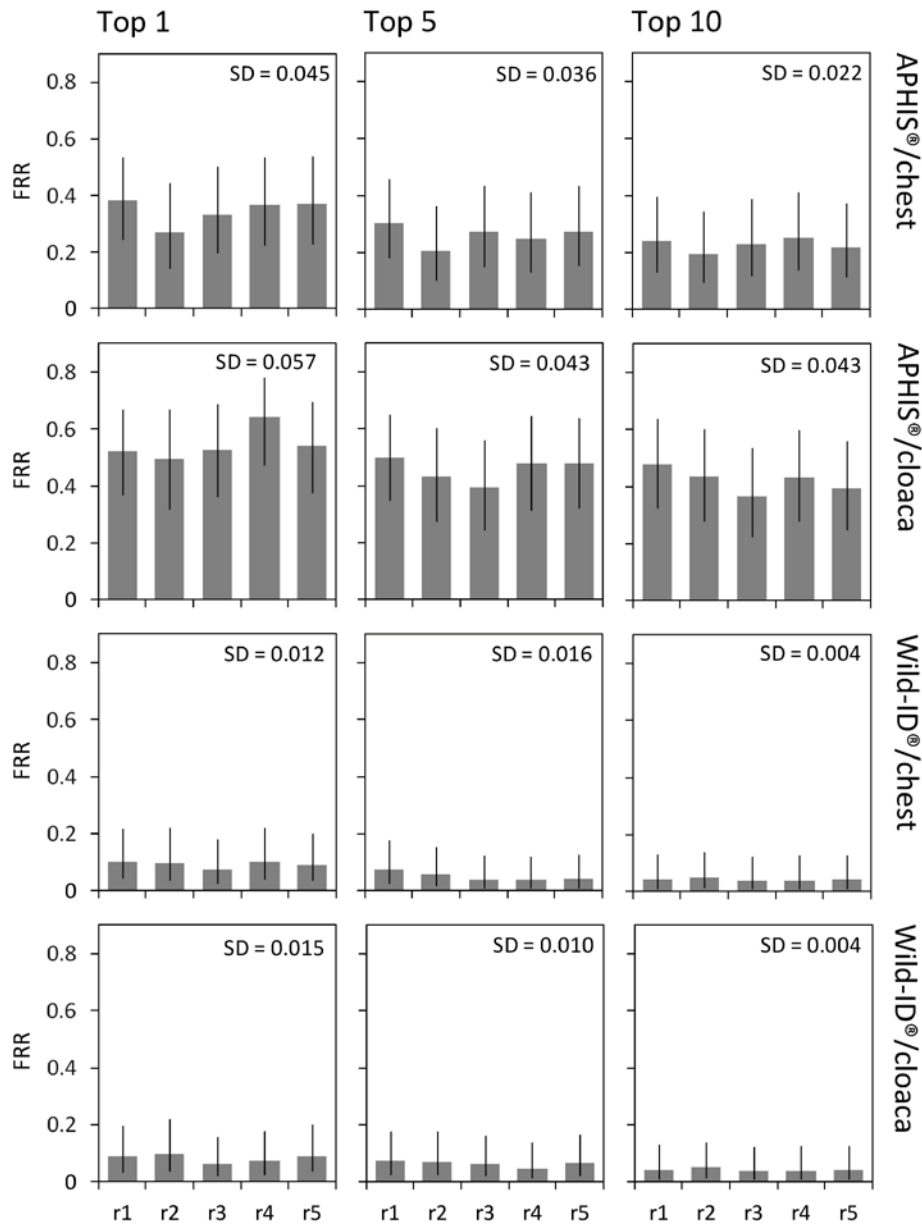
Figure 2. Software reliability: variability for each combination of software, body region and Top rank matches of the False Rejection Rate (FRR) across five replicates (r1–r5) of the matching of individual recapture pictures of the French cave salamander *Hydromantes strinatii*. Grey columns show the estimated mean FRR and black bars their 95% CI.

with poorly-standardised conditions. Even in our study, software effectiveness could have been improved further with even higher quality pictures produced by (i) holding individuals manually or using a shutter with a hole without glass (i.e. allowing a direct view on a selected body region) in order to avoid bubbles and impurities, (ii) improving lighting, and therefore the depth of field, for instance using a powerful flash, and (iii) holding the camera still with a tripod. Also, cropping pictures is supposed to improve software effectiveness. By comparing here FRRs obtained with uncropped pictures (Wild-ID) vs pictures where we restricted the searching algorithm to the focus area by generating a set of landmarks (APHIS), we potentially skewed our results in favour of APHIS. However, the little improvement cropping pictures provided in Wild-ID picture ranking had no effect on the computing of FRR values. If it had,

it only had strengthened the dominance of Wild-ID over APHIS.

Using APHIS with the ITM procedure on chest pictures of a salamander, Moya et al. (2015) detected 100 % of recaptures (i.e. FRR = 0, *n* = 305). In contrast, we obtained with APHIS poor FRR values (i.e. lower effectiveness), which were in addition highly variable (i.e. lower reliability) (Fig. 2). This is probably due to our low level of standardisation (e.g. body distortion, variable lighting), for which the software cannot fully compensate despite pre-processing of images, combined with human error since pre-processing is executed by hand during each replicate by positioning two new reference points on the same pictures.

Wild-ID was as effective with pictures of the chest as with those of cloaca. On the other hand, APHIS appeared to be less effective with cloaca pictures than with chest

Table 2. Review of FRRs obtained with Wild-ID when monitoring amphibia in different studies. PAS = "point and shoot" camera; DSLR = "digital single lens reflex" camera; CC = Camcorder

High-score matching range	Species	Pattern	Type of camera	FRR (sample size)	References
Top 100	*Eurycea tonkawae*	Dorsal head	DSLR	0.008 (1367)	Bendik et al., 2013
			PAS	0.159 (965)	
Top 20	*Anaxyrus baxteri*	Dorsal side	DSLR	0.200 (822)	Morisson et al., 2016
			PAS	0.470 (822)	
	Melanophryniscus cambaraensis	Ventral side	PAS	0.091 (492)	Coarsi et al., 2012
	Melanophryniscus montevidensis	Ventral side	PAS	0.100 (410)	Elgue et al., 2014
	Salamandrina perspicillata	Ventral side	?	0.650 (760)	Romiti et al., 2016
Top 10	*Hydromantes strinatii*	Cloaca region	PAS	0.042 (253)	This study
		Chest region	PAS	0.042 (253)	
	Ichthyosaura alpestris	Body (left) side	DSLR	♀: 0.001 (721) ♂: 0.000 (517)	Mettouris et al., 2016
	Lissotriton vulgaris	Ventral side	DSLR	♀: 0.031 (125) ♂: 0.025 (77)	
	Salamandra salamandra + S. infraimmaculata	Ventral side	PAS - CC	0.732 (500) 0.770 (2000) 0.774 (2197)	Matthé et al., 2017
	Triturus carnifex + T. cristatus	Ventral side	PAS - CC	0.146 (500) 0.161 (2000) 0.169 (4000) 0.186 (7000) 0.188 (7458)	Matthé et al., 2017
	Ambystoma opacum	Ventral side	PAS - CC	0.177 (500) 0.217 (2000) 0.264 (4000) 0.304 (7000) 0.341 (12488)	Matthé et al., 2017
	Bombina variegata	Ventral side	PAS - CC	0.027 (500) 0.033 (2000) 0.036 (4000) 0.036 (4063)	Matthé et al., 2017
Top 1	*Hydromantes strinatii*	Cloaca region	PAS	0.082 (253)	This study
		Chest region	PAS	0.093 (253)	
	Ichthyosaura alpestris	Body (left) side	DSLR	♀: 0.017 (710) ♂: 0.006 (514)	Mettouris et al., 2016
	Lissotriton vulgaris	Ventral side	DSLR	♀: 0.186 (105) ♂: 0.076 (73)	Mettouris et al., 2016
	Salamandra salamandra + S. infraimmaculata	Ventral side	PAS - CC	0.863 (500) 0.882 (2000) 0.884 (2197)	Matthé et al., 2017
	Triturus carnifex + T. cristatus	Ventral side	PAS - CC	0.274 (500) 0.300 (2000) 0.324 (4000) 0.352 (7000) 0.355 (7458)	Matthé et al., 2017
	Ambystoma opacum	Ventral side	PAS - CC	0.350 (500) 0.449 (2000) 0.505 (4000) 0.553 (7000) 0.604 (12488)	Matthé et al., 2017
	Bombina variegata	Ventral side	PAS - CC	0.047 (500) 0.059 (2000) 0.068 (4000) 0.068 (4063)	Matthé et al., 2017

Table 3. Population size estimated from a simulated French cave salamander population of 600 individuals. 1000 datasets with randomly varying detection probabilities over six closed capture events (mean detection probability of 0.1) were simulated. False Rejection Rates (FRRs) obtained from Table 1 (chest Wild-ID Top 10 and Top 1 matches) were applied to simulated datasets. Population sizes were estimated applying standard closed capture model with a time effect on detection probability.

Dataset	FRR	Mean estimated population size [95 % CI]
Control	0	588 [500-687]
Top 10	0.042	618 [522-712]
Top 1	0.093	655 [554-755]
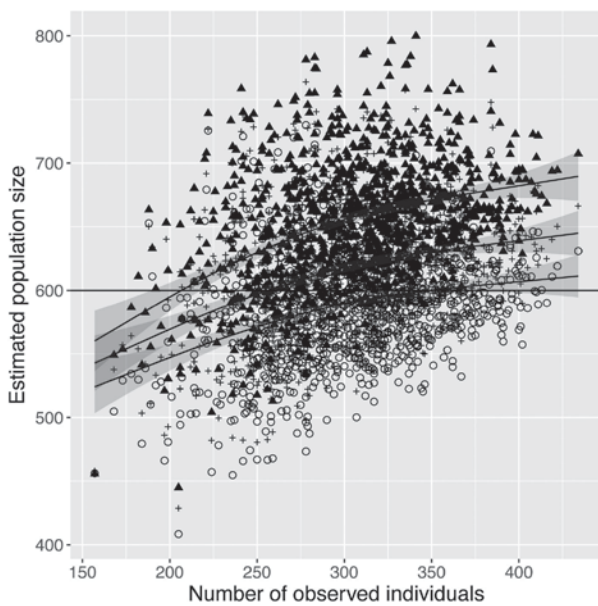
Figure 3. Impact of false rejections on estimation of abundance of a simulated population of 600 French cave salamanders. Capture-recapture data were collected over six closed capture events with mean detection probability of 0.1. Estimated population size from simulated data with no error are presented with open circles, from simulated data with False Rejection Rate (FRR) corresponding to the use of Top 10 Wild-ID on the chest are presented with crosses and with FRR corresponding to the use of Top1 Wild-ID on the chest are presented with black triangles. A regression spline smoother with 95% confidence intervals was added for each group of data to help with visual interpretation. The smoother was fitted using the mgcv package (Wood, 2006) and explains respectively 10%, 12% and 17% of the variation in the estimated population sizes for the Control, Top10 and Top1 datasets.

pictures. When flattened against the glass, the soft cloaca can vary in shape, therefore modifying patterns. The higher sensitivity of APHIS to distortion is most probably once again due to the pre-processing conducted by hand. Pictures of the less mobile and more stable body regions such as the chest produced results at least as good as the cloaca pictures in our study, so we suggest using pictures from such stable body regions with all software. Though I3S straighten software copes with body contortion, it requires a manual pre-processing approach (Den Hartog & Reijns, 2015), which is time consuming.

Consequences for the estimation of population size
One assumption necessary for valid inference from capture-mark-recapture studies is that marks are correctly read (Lindberg, 2012). In case of non-invasive sampling, errors in the identification of individuals may lead to bias in estimation of demographic parameters (Creel et al., 2003). Here we showed that considering first Wild-ID proposals (Top 1 high-score matches) leads to an overestimation of population size of 9%. Underestimation of population size without FRR was due to the low detection probability found in this species (<0.1, see Methods section) and applied to the simulated dataset. Indeed, datasets created from simulations that implied lower number of recaptures induced higher uncertainty and underestimation of the population size (Fig. 3). As a consequence, we found wide 95 % confidence intervals in the estimated population size (Table 3).

A 9% overestimation of population size is however twenty times less than the overestimation of population size found by Morrison et al. (2016) using Wild-ID. This is consistent with the much higher FRR (0.47 for Top 20) found by Morrison et al. (2016) on Wyoming toad (*Anaxyrus baxteri*) than those found in this study. In order to limit bias in estimation of population size when individuals are misidentified, closed capture models incorporating misidentification as implemented in the programme MARK may be used but, as a consequence in such case, confidence intervals would become much wider (Lukacs & Burnham, 2005). When a long-term monitoring programme is launched, a further important point to consider is the possibility that chromatophore patterns vary over time in amphibia (Church et al., 2007; Drechsler et al., 2015; Balogova & Kyselova, 2016).

Management implications
The assessment of bias in individual recognition is an essential step when software is used to estimate the demographic parameters, especially in a threatened species. Indeed, these values can greatly influence the decisions regarding conservation strategies. To limit this bias without spending too much time reviewing all available pictures, especially with huge sample sizes (i.e. *n* > 1000), we recommend considering the Top 10 high-score matches proposed by Wild-ID, at least for *H. strinatii* and most certainly for several other *urodela* species that have a similar pigment pattern. Our study suggests that in the conditions described here (low detection probability), overestimation of the population size is 3 %, and true population size falls within the 95% confidence interval.

Although Wild-ID was initially developed for large terrestrial mammals, we show that the software is able to cope effectively with a small sized amphibian with a complex chromatophore pattern. We have also shown that it is effective in field conditions and without pre-processing treatment of pictures. Therefore, it can work

well even with limited equipment and time. This also means stress to the animals and the cost of the study may be reduced using Wild-ID than other software – two aspects which are nowadays of primary concern for reasons of ethics and efficiency. The low FRR values obtained through the present study in *H. strinatii* provide encouraging prospects for using Wild-ID to identify and monitor species with complex pigment patterns. Even if the FRR may increase with the size of the dataset, this rate can be reduced by increasing the quality of the pictures (Gamble et al., 2008; Matthé et al., 2017).

ACKNOWLEDGMENTS

REFERENCES

Arntzen, J.W., Goudie, I.B.J., Halley, J. & Jehle, R. (2004). Cost comparison of marking techniques in long-term population studies: PIT- tags versus pattern maps. *Amphibia-Reptilia* 25, 305–315.

Balogova, M. & Kyselova, M. (2016). Changes in dorsal spot pattern in adult *Salamandra salamandra* (Linnaeus, 1758). *Herpetozoa* 28, 167–171.

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7. Available at http://CRAN.R-project.org/package=lme4.

Bendik, N., Morrison, T.A., Gluesenkamp, A.G., Sanders, M.S. & O'Donnell, L., (2013). Computer-assisted photo identification outperforms visible implant elastomers in an endangered salamander, *Eurycea tonkawae*. *PLoS One* 8, 1–8. e59424.

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, 813–822.

Caorsi, V.Z., Santos, R.R. & Grant, T., (2012). Clip or Snap? An evaluation of toe-clipping and photo-identification methods for identifying individual Southern red-bellied toads, *Melanophryniscus cambaraensis*. *South American Journal of Herpetology* 7, 79–84.

Chao, A. (1989). Estimating population size for sparse data in capture-recapture experiments. *Biometrics* 45, 427–438.

Church, D.R., Bailey, L.L.,Wilbur, H.M., Kendall, W.L. & Hines, J.E. (2007). Iteroparity in the variable environment of the salamander *Ambystoma tigrinum*. *Ecology* 88, 891–903.

Crall, J., Stewart, C., Berger-Wolf, T.Y., Rubenstein, D. & Sundaresan, S.R. (2013). Hotspotter—patterned species instance recognition. In 2013 *IEEE Workshop on Application of Computer Vision*, 230–237.

Creel S., Spong, G., Sands, J.L., Rotella, J., Zeigle, J., Joe, L., Murphy, K.M. & Smith, D. (2003). Population size estimation in Yellowstone wolves with error-prone noninvasive microsatellite genotypes. *Molecular Ecology* 12, 2003–2009.

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, 315–325.

Den Hartog, J.E. & Reijns, R. (2015). Interactive individual identification system (I3S), Straighten. Version 1.0. Reijns Free Software Foundation Inc, Boston. Available at http://www.reijns.com/i3s/download/I3S_download.html.

Drechsler, A., Helling, T. & Steinfartz, S. (2015). Genetic fingerprinting proves cross-correlated automatic photo-identification of individuals as highly efficient in large capture–mark–recapture studies. *Ecology & Evolution* 5, 141–151.

Elgue, A., Peirera, G., Achaval-coppes, F. & Maneyro, R. (2014). Validity of photo-identification technique to analyze natural markings in *Melanophryniscus montevidensis* (Anura: Bufonidae). *Phyllomedusa* 13, 59–66.

Ferner, J.W. (2010). Measuring and marking post-metamorphic amphibians. In *Amphibian Ecology and Conservation: A Handbook of Techniques*, 123–141. Dodd, C.K. (ed). England: Oxford University Press.

Fox, J., Weisberg, S., Friendly, M., Hong, J., Andersen, R., Firth, D. & Taylor, S. (2016). Effect displays for linear, generalised linear, and other models. R package version 3.1-2. Available from: https://cran.r-project.org/web/packages/effects/effects.pdf

Fu, V.W.K., Karraker, N.E. & Dudgeon, D. (2013). Breeding Dynamics, Diet, and Body Condition of the Hong Kong Newt *Paramesotriton hongkongensis*. *Herpetological Monographs* 27, 1–22.

Gamble, L., Ravela, S. & McGarigal, K. (2008). Multi-scale features for identifying individuals in large biological databases: an application of pattern recognition technology to the marbled salamander *Ambystoma opacum*. *Journal of Applied Ecology* 45, 170–180.

Griffiths, R. A., Foster, J., Wilkinson, J.W. & Sewell, D. (2015). Science, statistics and surveys: a herpetological perspective. *Journal of Applied Ecology* 52, 1413–1417.

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological Monographs* 54, 187–211.

Jain, A.K. (2007). Biometric recognition. *Nature* 449, 38–40.

Kuznetsova, A., Brockhoff, P.B. & Christensen, R.H.B. (2017). Tests in linear mixed effects models. R package version 3.0-1. Available at https://cran.r-project.org/web/packages/lmerTest/lmerTest.pdf.

Lanza, B., Caputo, V., Nascetti, G. & Bullini, L. (1995). Morphologic and genetic studies on the European plethodontid salamanders: taxonomic inferences (genus *Hydromantes*). Museo Regionale di Scienze Naturali, Italia.

Lanza, B., Pastorelli, C., Laghi, P. & Cimmaruta, R. (2006). A review of systematics, taxonomy, genetics, biogeography and natural history of the genus *Speleomantes dubois*, 1984 (Amphibia Caudata Plethodontidae). Atti del Museo Civico di Storia Naturale di Trieste 52, Italia.

Lindberg, M.S. (2012). A review of designs for capture–mark–recapture studies in discrete time. *Journal of Ornithology* 152, 355–370.

Lowe, D. (2004). Distinctive image features from scale-invariant key-points. *International Journal of Computer Vision* 60, 91–110.

Lukacs, P.M. & Burnham, K.P. (2005). Estimating population size from DNA-based closed capture-recapture data incorporating genotyping error. *Journal of Wildlife Management* 69, 396–403.

Matthé, M., Schönbrodt, T. & Berger, G. (2008). Computergestützte Bildanalyse von auchfleckenmustern des Kammmolchs *Triturus cristatus*. *Zeitschrift für Feldherpetologie* 15, 89–94.

Matthé, M., Sannolo, M., Winiarski, K., Spitzen-van der Sluijs, A., Goedbloed, D., Steinfartz, S. & Stachow, U. (2017). Comparison of photo-matching algorithms commonly used for photographic capture–recapture studies. *Ecology & Evolution* 7, 5861–5872.

McCrea, R.S. & Morgan, B.J.T. (2014). Analysis of Capture-Recapture Data. Chapman and Hall/CRC Press, Florida, USA.

Mettouris, O., Megremis, G. & Giokas, S. (2016). A newt does not change its spots: Using pattern mapping for the Identification of individuals in large populations of newt species. *Ecological Research* 31, 483–489.

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, 454–463.

Morrison, T. A., Keinath, D., Estes-Zumpf, W., Crall, J. P. & Stewart, C.V. (2016). Individual Identification of the Endangered Wyoming Toad *Anaxyrus baxteri* and Implications for Monitoring Species Recovery. *Journal of Herpetology* 50, 44–49.

Moya, Ó., 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.

Nichols, J.D. (1992). Capture-recapture models. *BioScience* 42, 94–102.

Otis, D. L., Burnham, K.P., White, G.C & Anderson, D.R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* 62, 3–135.

Plummer, M. (2003). JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. In: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing* (DSC 2003), March 20–22. Vienna, Austria.

Renet, J. & Delauge, J. (2012). Vers la mise en place d'une stratégie conservatoire en faveur du Spéléomante de Strinati *Speleomantes strinatii* (Aellen, 1958) dans le sud-est de la France. *Nature de Provence* 1, 5–13.

R Development Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria, R Foundation for Statistical Computing. Available at http://www.R-project.org.

Romano, A., Basile, M. & Costa, A. (2018). Skewed sex ratio in a forest salamander: artefact of the different capture probabilities between sexes or actual ecological trait? *Amphibia-Reptilia* 39, 79–86.

Romano, A., Costa, A., Basile, M., Raimondi, R., Posillico, M., Scinti Roger, D., Crisci, A., Piraccini, R., Raia, P., Matteucci, G. & De Cinti, B. (2017). Conservation of salamanders in managed forests: methods and costs of monitoring abundance and habitat selection. *Forest Ecology & Management* 400, 12–18.

Romiti, F., Bissattini, A.M, Buono, V., Cifarelli, C., Rocca, F., Eniang, E.A., Akani, G.C. Luiselli, L., Superti, V., Carpaneto, G.P. & Vignoli, L. (2016). Photographic Identification Method (PIM) using natural body marks: a simple tool to make a long story short. *Zoologischer Anzeiger - A Journal of Comparative Zoology* 266, 136–147.

Sacchi, R., Scali, S., Mangiacotti, M., Sannolo, M. & Zuffi, M.A.L. (2016). Digital identification and analysis. In *Reptile Ecology and Conservation: A Handbook of Techniques*, 59–72. Dodd, Jr., C.K. (ed). Oxford: Oxford University Press.

Sannolo, M., Gatti, F., Mangiacotti, M., Scali, S. & Sacchi, R. (2016). Photo-identification in amphibian studies: a test of I3S Pattern. *Acta herpetologica* 11, 63–68.

Su, Y.-S. & Yajima, M. (2012). R2jags: A Package for Running jags from R. R Package Version 0.03-08, URL Http://CRAN.R-Project. Org/Package= R2jags.

Temple, H. J. & Cox, N. A. (2009). European Red List of Amphibians. Office for Official Publications of the European Communities. IUCN Publications Services, Luxembourg.

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 taurus*. *Journal of Applied Ecology* 44, 273–280.

Wengert, G.M. & Gabrial, M.W. (2006). Using chin spot patterns to identify individual mountain yellow-legged frogs. *Northwestern Naturalist* 87, 192.

Wood, S.N. (2006). Generalized additive models: an introduction with R. Chapman and Hall/CRC, London.

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, 3–9.