

## The Green Toad example: a comparison of pattern recognition software

Stephan BURGSTALLER<sup>1,2</sup>, Günter GOLLMANN<sup>2</sup> and Lukas LANDLER<sup>1</sup>

1. Institute of Zoology, University of Natural Resources and Life Sciences, 1180 Vienna, Gregor-Mendel-Straße 33/I, Austria.

2. University of Vienna, Department of Evolutionary Biology, 1090 Vienna, Althanstraße 14, Austria.

\* Corresponding author, L. Landler, E-mail: [lukas.landler@boku.ac.at](mailto:lukas.landler@boku.ac.at)

Received: 14. November 2020 / Accepted: 17. March 2021 / Available online: 25. March 2021 / Printed: June 2021

**Abstract.** Individual identification of animals is important for assessing the size and status of populations. Photo-based approaches, where animals are recognized by naturally occurring and visually identifiable features, such as color patterns, are cost-effective methods for this purpose. We compared five available programs (Wild-ID, Hotspotter, AmphIdent, APHIS, I3S Pattern+) for their power to semi-automatically identify dorsal patterns of the European green toad (*Bufo viridis*). We created a data set of 200 pictures of known identity, two pictures for each individual, and analyzed the percentage of correctly identified animals for each software. Furthermore, we employed a generalized linear mixed model to identify important factors contributing to correct identifications. We used these results to estimate the population size of our hypothetical population. The freely available HotSpotter application performed best, identifying close to 100 % of the photos correctly. Females had a higher chance of recognition, presumably because of sex-specific differences in the pattern contrast. Population estimates were close to the expected 100 for HotSpotter, but for the other applications population size was highly overestimated. Given the clarity of our results we strongly recommend the HotSpotter software, which is a highly efficient tool for individual pattern recognition.

**Key words:** Amphibia, color pattern, photo-identification software, population statistics, recapture study.

### Introduction

One of the most basic but also most important tasks in conservation biology is to assess the size and status of animal populations. Therefore, mark-recapture techniques in combination with population statistics are used to estimate demographic parameters. In many cases, such data are collected in long-term studies and allow prediction of population trajectories. Often, mark recapture techniques rely on the tagging of animals with more or less invasive methods ranging from toe clipping in amphibians (Kenyon et al. 2009), inserting PIT tags (Hare et al. 2007) to adding visual markings (Donnelly et al. 1994). Alternatively, genetic identification can be used as a non-invasive method (Woods et al. 1999).

The most cost-effective and non-invasive method, however, is to make use of naturally occurring body patterns (Coarsi et al. 2012, Bendik et al. 2013, Drechsler et al. 2015). Unfortunately, visually matching pictures of such patterns can be very time consuming and, especially for long-term studies, decreases efficiency (Arntzen et al. 2004). Therefore, a number of pattern-matching algorithms have been developed which can cut time of photo comparisons, but in many instances, their application can also lead to decreased accuracy (Coarsi et al. 2012, Elgue et al. 2014, Morrison et al. 2016). Decreased accuracy, in turn, can cause considerable errors when estimating demographic parameters (Morrison et al. 2011, Winiarski & McGarigal 2016). Consequently, deciding on the right algorithm to use for the respective data set is crucial.

Differences between data sets could include distance to the animals (c.f. Bolger et al. 2012 and Kim et al. 2017), background of the photos, and their general quality (camera type, resolution, angle, blurriness (Patel & Das 2020)). As picture quality and attributes may vary strongly, scientists often prepare the photos before comparing the patterns, for example by cropping the pictures to only show the region of interest (ROI). This can be a quite elaborate task, although it is arguably less time consuming than comparing pictures visually. Matthé et al. (2017) compared the performance of four

pattern recognition programs with various amphibian species with high contrast patterns. Bardier et al. (2020) used two of these programs to compare their performance to that of visual recognition on *Ceratophrys stolzmanni*. In our case study we extended the selection of software packages and used different picture preparations, where applicable. Furthermore, we chose the green toad as an example of the programs' effectiveness on species with patterns of highly variable and changing contrast. We compared five different pattern recognition programs, four of which are freely available (exception: AmphIdent). Our results can help population ecologists to gain efficiency when comparing a variety of animal patterns for individual identification.

### Material and Methods

We used pictures from a long-term capture-recapture study of green toads (*Bufo viridis*) at the Rudolf-Bednar-Park in Vienna (48° 13' 35" N, 16° 23' 47" E). Green toads are suitable for automated pattern matching as they have striking dorsal patterns, although the contrast of this pattern can change over time (Goubeaud 1931). Pictures of the dorsal patterns of green toads were taken in the field with a variety of cameras. The photo background was mostly uniformly white or grey (sheet, cup etc.). A few photos were taken on patchy background (e.g. gravel) or contain parts of a researcher's hand. We classified each camera as one of three types: smart phone cameras (4.9–20.7 MP), compact cameras (8.1–24.3 MP) and reflex cameras (12.8–36.3 MP). For the purpose of this study, we selected 200 pictures of known animal IDs (two per ID, 100 'reference' and 100 'unknown') of varying quality and different sexes (females: 18, males: 58, unidentified: 24, this resembles naturally occurring ratios, for photos see: Burgstaller et al. 2021). Most matching pictures were taken in the same year, although the time interval between the pictures was over two years in some instances (median: 22.5 days; range: 2–770 days). We used five different software applications (Wild-ID (Bolger et al. 2012), HotSpotter (Crall et al. 2013), AmphIdent (application for fire bellied toads, pattern most similar to the green toad from all available applications) (Matthé 2018), I3S Pattern+ (Speed et al. 2007) and APHIS (Moya et al. 2015)) to compare the pictures. Furthermore, for Wild-ID and HotSpotter we used different cropping extents: a (whole animal), b (head) and c (ROI on head) (Figure 1). In Amph-

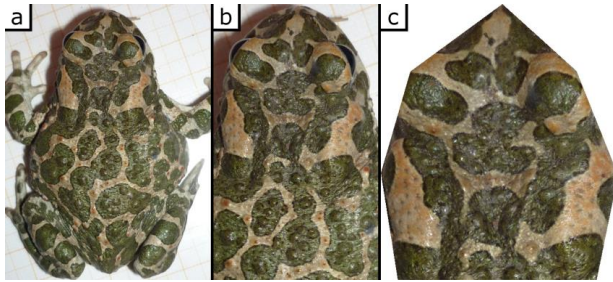


Figure 1. The three different cropping types used in the study for HotSpotter and Wild-ID: whole animal (a), head (b), region of interest (c).

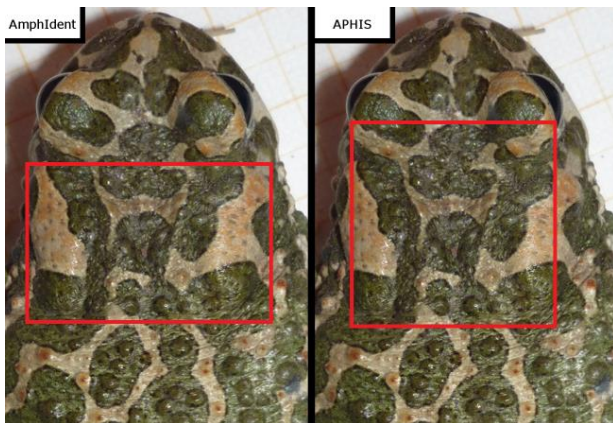


Figure 2. The shapes of the ROI for AmphIdent (left) and APHIS (right). The width to height ratio for APHIS is fixed at 1:1. We chose to locate the upper two corners of the ROI at the lower end of the visible eyeballs. For AmphIdent the ROI ratio can be freely chosen but the selected area will be stretched to a 10:7 ratio, so in order not to distort the original picture too much we selected an area with this approximate ratio. The ROI extends from the upper to the lower, as well as from the left to the right edges of the parotoid glands.

dent, APHIS and I3SP+ it is required to define a region of interest for every image after loading. For I3SP+ we defined a ROI similar to the shape of cropping type c. For AmphIdent and APHIS we had to adhere to the ROI selection of the respective software. These are squares and rectangles, respectively, which we placed in the same part of the head as in cropping type c (see Figure 2 for details). AmphIdent and I3SP+ have functions for manual input to improve the performance, we used them to the following extent. In I3SP+ pixels can be chosen to discern the pattern's fore- and background. Our time effort to discern fore- from background did never exceed three minutes. AmphIdent has the option to only compare subareas of the ROI. If comparing the whole ROI did not yield a true detection, this function was used up to two times.

Pictures were always compared in the same order. The picture comparisons followed the applications' manuals; however, the general procedure was the same throughout. We first added 100 'reference' pictures for each of the used individuals to the application's database, which simulates that we already encountered each of the individuals (once) before. We did this to ensure that each of the new pictures could find a match in the previous record. We then compared 100 pictures of 'unknown' ID to the application database. Usually, the user decides whether one of the first few automatically selected pictures is a match or not, but for the purpose of this analysis we only considered the picture ranked first. We recorded a true (1) or false (0) detection and bootstrapped a 95 % confidence interval for the mean detection success for each software using the function *boot.ci* in the package *boot* (Canty & Ripley 2020) in R (R Core Team

2020) (see supplementary files - available online).

For programs that performed well enough so that further automatization could be considered we created histograms to visualize the scores calculated by the program for each picture pair. We compared the matching scores (MS) of all correct picture pairs ( $n = 100$ ) to all non-matching scores (NMS) of the other picture pairs ( $n = 19,900$ ) and added a vertical threshold line at the position of the highest NMS.

To statistically analyze the performance of the different software we used a generalized linear mixed model (GLMM) with a binomial error distribution employing the function *glmmTMB* from the package *glmmTMB* (Brooks et al. 2017). As the response variable we used the level of detection (0/1) and predictors were the software application, sex of the animals, camera type of the first picture and the second picture as well as interaction between the camera types. In addition, we used individual as a random factor in the model. The use of GLMM allows to statistically test the influence of multiple factors on the performance of detection algorithms, extending the insights we gain from such a software comparison study.

We used the Petersen index (Chao et al. 2008) to estimate abundance assuming that we compared two sampling occasions (the first occasion was the set of the first 100 pictures of 100 individuals, the second occasion was the second set of 100 pictures). The true 'population size' therefore was 100. Raw data and R code to reproduce our analysis are available in the supplementary information.

## Results

From the evaluated applications HotSpotter performed the best with detection rates close to 100 %, with only little variation accounted for by image cropping for this software (Figure 3). AmphIdent, Wild-ID (c), and APHIS identified just over 60 % of the images correctly. Wild-ID (b), Wild-ID (a) and I3SP+ all identified less than 60 % of the images correctly. Cropping was important for Wild-ID, but not for HotSpotter. It should be noted that in all cases results would have been substantially better, if we had included more than just the first hit for an image.

Only HotSpotter performed well enough for considering a further automated application, where true matches can be obtained with using a matching score threshold. For this program our analysis suggested a matching score  $>2700$ ,

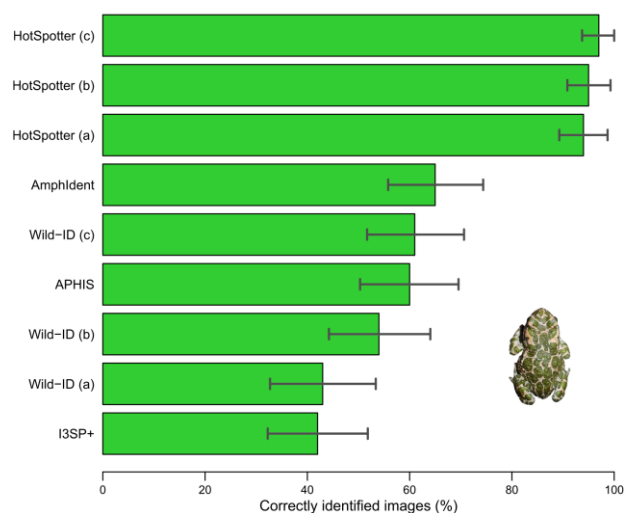


Figure 3. Results of comparison showing the percentage of correctly identified pictures in our data set. Error bars indicate bootstrapped confidence intervals.

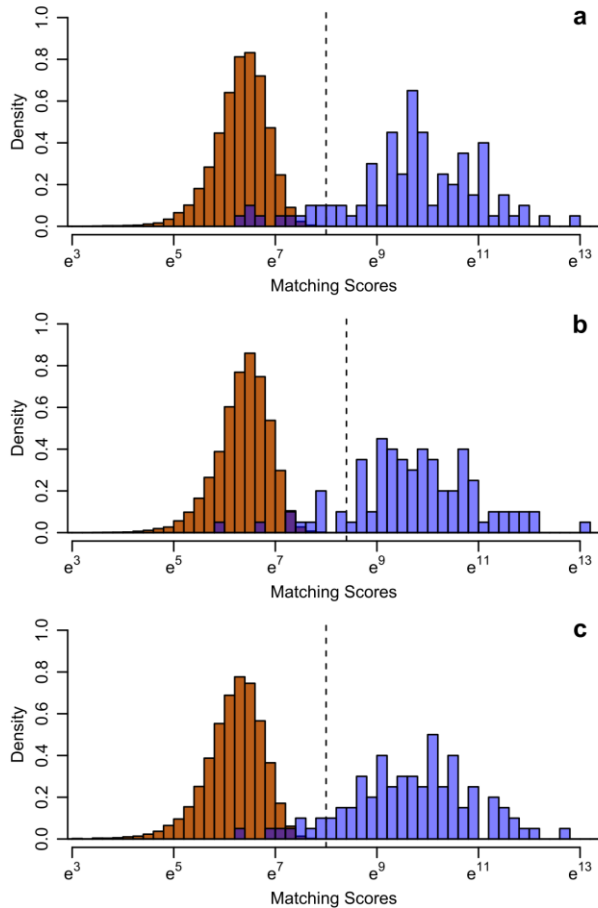


Figure 4. Histograms comparing the matching scores (blue) and non-matching scores (orange) calculated by the software HotSpotter for three cropping styles (a, b, c).

>3600, and >2500 for cropping types a, b, and c, respectively, as a suitable cut-off for true matches (see Figure 4). Pictures with lower scores may be manually double checked. Only 10 % (a, b) and 8 % (c) of MS overlapped with the NMS, thus 90 % of the matches could be successfully recognized in a fully automated process without double-checking.

Females achieved significantly higher individual recognition scores than males (Table 1). The camera type used for the pictures only showed weakly significant effects or no effects at all, no interaction effects between the camera used for the first and the second photo were detected (Table 1).

If we used solely the first hit from the respective software application and no additional manual comparison, the population sizes would be overestimated to different extents (Table 2). For the HotSpotter software, the estimated numbers are still close to the actual  $N$ , and only a few animals would be added. However, for I3SP+ the estimated size would be more than doubled, in comparison to the real population size.

## Discussion

We show that using automatic pattern recognition applications can be a highly efficient tool to identify individual animals. Our results resemble those obtained previously by Matthé et al. (2017), with the addition of the freely available

Table 1. ANOVA table showing the results for the GLMM.

	Chisq	df	p
Sex	27.56	2	<0.01
Software	117.26	8	<0.01
Cam 1	7.39	2	0.025
Cam 2	2.10	1	0.147
Cam 1 : Cam 2	4.61	2	0.1

Table 2: Population size estimates for the different applications.

Software	Peterson index estimate
HotSpotter (c)	103.09
HotSpotter (b)	105.26
HotSpotter (a)	106.38
AmphIdent	153.85
Wild-ID (c)	163.93
APHIS	166.67
Wild-ID (b)	185.19
Wild-ID (a)	232.56
I3SP+	238.10

HotSpotter (Crall et al. 2013) software to the comparison. This software easily outcompeted all the others tested. Using this software, we identified the majority of the individuals correctly, even with our stringent detection criteria. Applying them for all other tested applications yielded detection rates around 60 % and below. The significantly higher detection success for females is likely due to the higher contrast of the female dorsal pattern (Goubeaud 1931). Nonetheless, it highlights potential confounding factors when performing capture-recapture analyses based on pattern recognition. Differences in the detection probability between individuals and sexes could bias population estimations. Varying distributions of different color morphs (Zhelev et al. 2020) might also result in unequal detection rates and thus varying biases across populations. Knowing such potential biases (e.g. detection probability for a certain group) and/or error rates could be used to inform and improve population analyses (Ellis 2018). In contrast, our analysis indicates that switching cameras from one to another occasion might not be an important factor for successful identification.

The HotSpotter software has the advantage that it performed very well using full body ROIs, presumably because it contains a well-developed background subtraction algorithm. This might be especially important when working with large databases, as a greater number of reference points might be needed than a reduced-area ROI can provide. Notably, the size of the database does not seem to change the relative performance of picture detection algorithms (Matthé et al. 2017), but this was not tested for HotSpotter. The good performance of HotSpotter potentially enables further automatization in picture matching processes. Our results show that, depending on the cropping type, 90 % of matches or more may be automatically identified without false positives. HotSpotter has been successfully employed in recapture studies on various species. Nipko et al. (2020) compared camera trap photos of jaguars (*Panthera onca*) and ocelots (*Leopardus pardalis*) and recorded detection rates of 85–99 % when using high quality pictures. Patel & Das (2020) reported ~94 % detection rate on dorsal pictures of *Amolops formosus*, whereas Dunbar et al. (2021) compared head photos of

hawksbill turtles (*Eretmochelys imbricata*) with a detection rate of ~80 %. All these detection rates refer to matches ranked first. For animals lacking well-defined color patterns, however, these applications might not be an efficient solution for individual identification (Morrison et al. 2016). Focusing on other distinct features (e.g. arrangement of warts or the pores of the parotoid gland) might be a viable alternative in these cases, possibly demanding different cropping techniques.

Recently, it has been shown that HotSpotter can be used for identification of frogs that were photographed from a distance of up to 6 m (Patel & Das 2020). Therefore, apart from the general use in traditional population analyses, matching individuals from photos could also be used in citizen science projects. In cases where citizens are asked to take pictures of certain animals in a specific area, scientists might be able to determine the individual identity from such photos (Smith et al. 2018). The camera type showing only minor effects on the recognition performance further suggests a good applicability on photos collected by citizen scientists using a wide range of different camera models.

Based on our green toad example, we show that the freely available HotSpotter software constitutes a major improvement to other more widely used pattern recognition software, such as APHIS and Wild-ID.

**Acknowledgement.** We thank Christoph Leeb, Eileen Heyer, Christian Wappl and the students who participated in the field courses 'Populationsbiologie heimischer Amphibien' and 'Freilandpraktikum Amphibienökologie' at the University of Vienna from 2015 to 2019, who helped in collecting the pictures in ongoing green toad monitoring efforts in Vienna. Capture and handling of toads were approved by the municipality of Vienna (permit MA 22 – 984143-2015-5). We thank two anonymous reviewers for their insightful comments.

## 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.
- 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: 17-28.
- Bendik, N.F., Morrison, T.A., Gluesenkamp, A.G., Sanders, M.S., O'Donnell, L.J. (2013): Computer-assisted photo identification outperforms visible implant elastomers in an endangered salamander, *Eurycea tonkawae*. Hemmi J.M. (ed.). *PLoS One* 8: e59424.
- Bolger, D.T., Morrison, T.A., Vance, B., Lee, D., Farid, H. (2012): A computer-assisted system for photographic mark-recapture analysis: software for photographic mark-recapture. *Methods in Ecology and Evolution* 3: 813-822.
- Brooks, M.E., Kristensen, K., Van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., Bolker, B.M. (2017): glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9: 378.
- Burgstaller, S., Gollmann, G., Landler, L. (2021): Photos for "The Green Toad Example." <https://doi.org/10.5281/zenodo.4588525>
- Canty, A., Ripley, B. (2020): boot: bootstrap R (S-Plus) functions. R package version 1.3-11.
- Chao, A., Pan, H.-Y., Chiang, S.-C. (2008): The Petersen-Lincoln estimator and its extension to estimate the size of a shared population. *Biometrical Journal* 50: 957-970.
- Coarsi, 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.
- Crall, J.P., Stewart, C.V., Berger-Wolf, T.Y., Rubenstein, D.I., Sundaresan, S.R. (2013): HotSpotter - Patterned species instance recognition. pp. 230-237. In: 2013 IEEE workshop on applications of computer vision (WACV). Institute of Electrical and Electronics Engineers.
- Donnelly, M.A., Guyer, C., Juterbock, J.E., Alford, R.A. (1994): Techniques for marking amphibians. pp. 277-284. In: Heyer, W., Donnelly, M.A., McDiarmid, R., Hayek, L.A., Foster, M.S. (eds.), *Measuring and monitoring biological diversity, standard methods for amphibians*. Smithsonian Institution Press, Washington DC.
- 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 and Evolution* 5: 141-151.
- 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.
- Elgue, E., Pereira, G., Achaval-Coppes, F., Maneyro, R. (2014): Validity of photo-identification technique to analyze natural markings in *Melanophryniscus montevidensis* (Anura: Bufonidae). *Phyllomedusa: Journal of Herpetology* 13: 59-66.
- Ellis, A.R. (2018): Accounting for matching uncertainty in photographic identification studies of wild animals. Theses and Dissertations--Statistics, University of Kentucky Libraries. <https://doi.org/10.13023/ETD.2018.026>
- Goubeaud, W. (1931): Die histologischen Grundlagen von Farbkleid und Farbwechsel bei *Bufo viridis*. *Zeitschrift für Morphologie und Ökologie der Tiere* 21: 702-739.
- Hare, K.M., Hoare, J.M., Hitchmough, R.A. (2007): Investigating natural population dynamics of *Nautilinus manukanus* to inform conservation management of New Zealand's cryptic diurnal geckos. *Journal of Herpetology* 41: 81-93.
- Kenyon, N., Phillott, A.D., Alford, R.A. (2009): Evaluation of the photographic identification method (PIM) as a tool to identify adult *Litoria genimaculata* (Anura: Hylidae). *Herpetological Conservation and Biology* 4: 403-410.
- Kim, M.Y., Borzée, A., Kim, J.Y., Jang, Y. (2017): Treefrog lateral line as a mean of individual identification through visual and software assisted methodologies. *Journal of Ecology and Environment* 41: art.42.
- Matthé, M. (2018): AmphIdent. Available from: <<http://www.amphident.de>>.
- 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 and Evolution* 7: 5861-5872.
- Morrison, T.A., Yoshizaki, J., Nichols, J.D., Bolger, D.T. (2011): Estimating survival in photographic capture-recapture studies: overcoming misidentification error: unbiased survival estimation in photograph-ID. *Methods in Ecology and Evolution* 2: 454-463.
- Morrison, T.A., Keinath, D., Estes-Zumpff, 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.
- 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: 424-433.
- Patel, N.G., Das, A. (2020): Shot the spots: a reliable field method for individual identification of *Amolops formosus* (Anura, Ranidae). *Herpetozoa* 33: 7-15.
- R Core Team (2020): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: <<https://www.r-project.org/>>.
- Smith, S., Young, T., Skydmore, D. (2018): Natterjack toads (*Epidalea calamita*) using comparisons of dorsal features through citizen science. *Herpetological Journal* 28: 31-38.
- Speed, C.W., Meehan, M.G., Bradshaw, C.J. (2007): Spot the match – wildlife photo-identification using information theory. *Frontiers in Zoology* 4: art.2.
- Winiarski, K.J., McGarigal, K. (2016): Effects of photo and genotype-based misidentification error on estimates of survival, detection and state transition using multistate survival models. *PLOS ONE* 11: e0145640.
- Woods, J.G., Paetkau, D., Lewis, D., McLellan, B.N., Proctor, M., Strobeck, C. (1999): Genetic tagging of free-ranging black and brown bears. *Wildlife Society Bulletin* 27: 616-627.
- Zhelev, Z., Mollov, I., Tsonev, S. (2020): Body size and color polymorphism in *Bufoles viridis* complex (Anura: Bufonidae) inhabiting two semi-natural areas in Plovdiv City, Bulgaria. *North-Western Journal of Zoology* 16(2): 191-196.