 DOI: 10.13157/arla.72.2.2025.fo1

Forum

The **Forum** section of *Ardeola* publishes papers whose main topic, contents and/or format differ from the normal articles and short notes published by the journal. Its main aim is to serve as a lighter channel for discussion and constructive criticism on papers or reseach lines published either in *Ardeola* or elsewhere, as well as to stimulate the publication of new ideas and short revisions on current ornithological topics.

***Forum** es una sección que pretende servir para la publicación de trabajos de temática, contenido o formato diferentes a los de los artículos y notas breves que se publican en* Ardeola. *Su principal objetivo es facilitar la discusión y la crítica constructiva sobre trabajos o temas de investigación publicados en* Ardeola *u otras revistas, así como estimular la presentación de ideas nuevas y revisiones sobre temas ornitológicos actuales.*

BIRDNET CONFIDENCE SCORES DECREASE WITH BIRD DISTANCE FROM THE RECORDER: REVISITING PÉREZ-GRANADOS (2023)

LAS PUNTUACIONES DE CONFIANZA DE BIRDNET DECRECEN CON LA DISTANCIA DEL AVE AL GRABADOR: REVISITANDO A PÉREZ-GRANADOS (2023)

Cristian Pérez-Granados[1]

Summary.—BirdNET is a machine learning tool capable of identifying 6,500 bird species worldwide. It is available on different platforms but with varying versions and algorithms, which may lead to misleading results. For this Forum, I reanalysed recordings from Pérez-Granados (2023) using BirdNET-Analyzer, the research-focused platform of BirdNET, instead of the demo platform (BirdNET-Api) used in the original study. BirdNET-Analyzer outperformed BirdNET-Api by detecting 28% more vocalisations and maintaining high detection probabilities over greater distances (75 metres vs. 50 metres). Unlike in the original study, there were significant differences in detection probabilities among

[1] Biodiversity Management and Conservation Programme, Forest Science and Technology Center of Catalonia (CTFC), 25280 Lleida, Spain.

Corresponding author: cristian.perez@ctfc.cat

recorder models and between the species more often detected. BirdNET-Analyzer confidence scores also decreased significantly with increasing distances from the bird to the recorder, and varied between species, and recorders – relationships not observed in the original study. Unlike in the original study, significant differences were found in detection probabilities among recorder models and in which species were more frequently detected. These findings may offer insights into future passive acoustic bird surveys using BirdNET, highlighting a trade-off between precision and detection radius and aiding in the modelling of detection radii for improved bird density estimation and comparability between studies. Overall, the variability in confidence scores across distances, recorders and species, emphasises the need for further research to optimise large-scale monitoring programmes using BirdNET, which often require identifying optimal species-specific confidence score thresholds. Finally, this Forum highlights the importance of preferring BirdNET-Analyzer for research purposes to avoid misleading interpretations. — Pérez-Granados, C. (2025). BirdNET confidence scores decrease with bird distance from the recorder: revisiting Pérez-Granados (2023). *Ardeola*, 72: 149-159.

Keywords: convolutional neural network, detection radius, machine learning, model confidence, threshold.

RESUMEN. — BirdNET es una herramienta de aprendizaje automático capaz de identificar 6.500 especies de aves en todo el mundo. BirdNET está disponible en diferentes plataformas, pero con versiones y algoritmos diferentes, lo que puede llevar a resultados confusos. En este *Forum*, volví a analizar las grabaciones de Pérez-Granados (2023) utilizando BirdNET-Analyzer, la plataforma de BirdNET enfocada en la investigación, en lugar de la plataforma de demostración (BirdNET-Api) utilizada en el estudio original. BirdNET-Analyzer superó a BirdNET-Api al detectar un 28 % más de vocalizaciones y mantener altas probabilidades de detección a mayores distancias (75 metros frente a 50 metros). A diferencia del estudio original, hubo diferencias significativas en las probabilidades de detección entre modelos de grabadoras y entre las especies detectadas con más frecuencia. Los puntajes de confianza de BirdNET-Analyzer también variaron: decrecieron significativamente según aumentaba la distancia del ave al grabador, y varía entre las diferentes especies y los tipos de grabadoras empleadas, relaciones que no se observaron en el estudio original. Estos hallazgos pueden proporcionar información valiosa para futuros estudios pasivos de monitoreo acústico de aves utilizando BirdNET, al poner de manifiesto que existe un compromiso entre la precisión y el radio de detección, y podría ayudar en el modelado de radios de detección para mejorar la estimación de densidad de aves y la comparabilidad entre estudios. En general, la variabilidad en los puntajes de confianza entre distancias, grabadores, y especies, enfatiza la necesidad de realizar más investigaciones para optimizar los programas de monitoreo a gran escala que usan BirdNET, los cuales a menudo requieren identificar umbrales óptimos de valores de confianza para cada especie. Finalmente, este *Forum* destaca la importancia de utilizar BirdNET-Analyzer para fines de investigación a fin de evitar interpretaciones erróneas. — Pérez-Granados, C. (2025). Las puntuaciones de confianza de BirdNET decrecen con la distancia del ave al grabador: revisitando a Pérez-Granados (2023). *Ardeola*, 72: 149-159.

Palabras clave: aprendizaje automático, confianza del modelo, radio de detección, red neuronal convolucional, umbral.

The advancements in automated and non-invasive techniques, combined with the development of machine learning tools, have transformed biodiversity monitoring methods (Lahoz-Monfort & Magrath, 2021; Kahl *et al*., 2021; Stowell, 2022). Among these advancements is BirdNET, a user-friendly machine learning tool able to identify, on sound recordings, the vocalisations of over 6,500 bird species worldwide (Kahl *et al*., 2021, reviewed by Pérez-Granados 2023a). BirdNET processes audio recordings by dividing them into three-second segments and for each segment can provide a multispecies prediction (Kahl *et al*., 2021). BirdNET predictions are accompanied by a quantitative confidence score ranging from 0 to 1, which indicates the model's certainty that a given prediction has been correctly identified (Wood & Kahl, 2024). Users can set a desired confidence score threshold to filter BirdNET output and include only predictions exceeding a specified probability of being correct (e.g. Bota *et al*., 2023; Singer *et al*., 2024). However, much further research is needed to optimise the use of its confidence scores (Wood & Kahl, 2024) and to evaluate how these scores vary with species and different contexts (Funosas *et al*., 2024).

BirdNET can be utilised through various platforms, including: i) a mobile application (Wood *et al*., 2022), ii) a web-based platform (BirdNET-API, see Pérez-Granados, 2023b), iii) through BirdNET-Analyzer, either via a graphical user interface (Bota *et al*., 2023) or Python (Manzano *et al*., 2022), and iv) an audio analyses software (Raven Pro). However, the algorithms and mechanisms implemented across the various platforms often differ, with only BirdNET-Analyzer being regularly updated and containing the most updated models (available at https://github.com/kahst/BirdNET-Analyzer). This inconsistency may hinder direct comparisons between studies using different platforms or BirdNET versions.

Pérez-Granados (2023b) investigated whether BirdNET performance varied with the distance of bird vocalisations from the recording equipment (hereon recorder) and between species. Overall, he found a significant decrease of BirdNET performance at increasing distances and large differences among species. Contrary to expectations, no relationship was found between BirdNET confidence scores for a given prediction and the distance at which the vocalisation was recorded. Pérez-Granados (2023b) suggested that the absence of this relationship was likely due to that "*over 90% of the identified vocalisations had a confidence score of 1*", making this parameter ineffective for inferring the distance of a bird from the recorder. However, such a high proportion of predictions with confidence scores equal to 1 are likely the result of processing the recordings using the web interface of BirdNET (see Pérez-Granados, 2023b), which was designed as a demo platform and does not contain the most updated algorithms of BirdNET (https://birdnet.cornell.edu/api/). Indeed, prior research has demonstrated that up to 80% of BirdNET predictions yield confidence scores below 0.5 (Wood *et al*., 2021; Bota *et al*., 2023). Overall, these findings suggest that the confidence scores reported through the web interface are not representative of those obtained using other BirdNET platforms, such as BirdNET-Analyzer, the platform with the most updated models.

In this Forum, the author re-analyzed the recordings collected by Pérez-Granados (2023b) using the latest version of BirdNET-Analyzer (v2.4 - v2, updated in January 2024). This Forum aims to discuss whether the results presented by Pérez-Granados (2023b) were misinterpreted owing to the BirdNET platform used. The final goal is to promote good research using BirdNET and specifically to assess whether there is a relationship between BirdNET confidence scores for a given prediction and the distance at which the vocalisation was broadcast. This study also aims to evaluate whether such relationships vary between recorder models and target species. For ease of reading hereon the term BirdNET-Api is used when referring to the BirdNET demo platform used by Pérez-Granados (2023b) and the term BirdNET-Analyzer when referring to the latest and more developed version of BirdNET.

The scanned recordings were the same ones obtained by Pérez-Granados (2023b) in "Pinada Villamontes" (San Vicent del

Raspeig, Alicante, eastern Spain, 38.26° N, 0.31° W). This is a flat, homogeneous and semi-open area with a low density of pines. Two Song Meter Micro and two Song Meter Mini devices were attached to a 70-90cm wooden stick and left in the same position (below forest canopy, with no direct blocking of sound, and with microphones oriented towards the loudspeaker) during the experiment. Recordings were collected using a sampling rate of 22.050 kHz, +18 dB gain, and 16-bit ".wav" format. A standardised playback, consisting of ten vocalisations each of the Hooded Warbler *Setophaga citrina*, Grasshopper Sparrow *Ammodramus savannarum*, and Gray Vireo *Vireo vicinior,* was broadcasted from seven different distances (10, 25, 50, 75, 100, 125, and 150 metres). The loudspeaker equipment consisted of a digital Samsung mp3 coupled with a RadioShack amplifier (frequency response 100 Hz-10 KHz, distortion 1 KHz < 2%), emitted with a volume around 76 dB (emulating normal singing volume of birds and measured with a Volcraft SL 400 sound level meter at 1m from the loudspeaker). Full details regarding the study area, recorders employed, vocalisations included in the playback, and loudspeaker used are provided in Pérez-Granados (2023b).

The recordings were analyzed using the latest version of BirdNET-Analyzer (version 2.4 – v2, Kahl *et al*., 2021) via the graphical user interface. To replicate the analyses performed by Pérez-Granados (2023b), BirdNET-Analyzer was configured with the minimum allowable confidence score threshold (0.01), the default sensitivity parameter (1.0), and no overlap for prediction segments (default setting). BirdNET-Analyzer was set to exclusively report detections of the three target species (Manzano-Rubio *et al*., 2022). The author repeated the same procedure as in Pérez-Granados (2023b) to assess whether BirdNET-Analyzer performance varies according to distances, recorders and species. Therefore, a logistic regression model was fitted using the glm function in R 4.4.2 (R Development Core Team, 2024). Goodness-of-fit of the logistic regression model was evaluated using McFadden's pseudo *r*-squared (R^2) with the "pscl" package (Jackman, 2024). The correct identification of a bird vocalisation (identified/not-identified) by BirdNET-Analyzer was considered as dependent variable, while distance (seven levels), recorder ID (four levels) and species (three levels) were considered as response variables.

A Tukey post-hoc comparison was run to test for differences among the levels of significant variables by using the "multcomp" package (Hothorn, 2008). To assess how BirdNET-Analyzer's confidence scores varied across distances, species and recorders, the author fitted a beta regression model with the "betareg" package (Zeileis *et al*., 2023). Beta regression models are specifically designed for continuous outcomes within the 0-1 interval. BirdNET-Analyzer's confidence scores were treated as the dependent variable, while distance (seven levels), recorder ID (four levels), and species (three levels) were included as explanatory variables. Post-hoc comparisons using the "emmeans" package (Lenth, 2025) were performed to evaluate pairwise differences between the levels of the significant factors. Goodness-of-fit of the beta regression model was evaluated using Cox and Snell's pseudo *r*-squared (R^2) estimated with the "performance" package (Lüdecke, 2021). In both models, the interaction between distance and recorder ID was initially included. However, all interaction terms had p-values greater than 0.95 in the logistic regression, and only three out of 17 interaction terms were significant ($p > 0.05$) in the beta regression model. In both cases, including the interaction did not improve model performance (lower AIC for models without interactions) and signs of overfitting were observed in the beta regression models when including the interaction. Therefore, the interaction between distance and recorder ID was not considered retained in the final models. The author also tested for multicollinearity between all variables in both models using the Generalized Variance Inflation Factor (GVIF) approach with the "car" package (Fox & Weisberg, 2019). The results showed no issues with multicollinearity in any model, with GVIF values < 1.3 in all cases.

BirdNET-Analyzer detected 641 bird vocalisations, a higher number compared to the 499 bird vocalisations detected by BirdNET-Api (Pérez-Granados, 2023b), representing a 28% increase in detections. Consistent with the prior research with BirdNET-Api, significant differences were observed in the probability of detecting bird vocalisations using BirdNET-Analyzer across distances and target species ($p < 0.001$ in both cases, see Supplementary material, Table S1, for statistical summaries and Supplementary material, Table S2, for post-hoc results). The probability of detecting bird vocalisations with BirdNET-Analyzer was similar across the first four distance categories (10m to 75m), but significantly higher than at distances beyond 75m (Figure 1a; Table S1, S2). In contrast, Pérez-Granados (2023b) identified just the first three categories (10m to 50m) as the radius beyond which BirdNET-Api's performance declined significantly. The larger radius of high detection probability observed with BirdNET-Analyzer likely corresponds with the greater number of vocalisations detected by this more advanced platform. The Tukey test revealed that BirdNET-Analyzer had a significantly higher probability ($p > 0.001$) of detecting Hooded Warbler vocalisations compared to those of the Gray Vireo and Grasshopper Sparrow (Figure 1b, Table S1, S2). In contrast, BirdNET-Api identified the Grasshopper Sparrow and Hooded Warbler with similar probabilities, both significantly higher than the Gray Vireo. While BirdNET-Api detected no significant differences in the probability of detecting bird vocalisations between recorders (Pérez-Granados, 2023b), BirdNET-Analyzer identified significant variations among recorder models. Specifically, the probability of detection was significantly higher ($p < 0.001$) for both Song Meter Mini recorders (174 and 176 vocalisations detected per unit) compared to the Song Meter Micro recorders (145 and 146 vocalisations detected per unit; Figure 1c, Table S1, S2). This finding is consistent with the superior performance of the Song Meter Mini microphones, which are more sensitive and external, whereas the Song Meter Micro devices rely on built-in microphones, potentially explaining their lower performance (see Darras *et al*., 2020).

Unlike the results obtained with BirdNET-Api (Pérez-Granados, 2023b), this study found that BirdNET-Analyzer's confidence scores –reflecting the model's certainty of correct identification (Wood & Kahl, 2024)– significantly decreased with increasing bird distance from the recorder (Figure 2a) and varied between recorders and target species ($p < 0.001$ in the three cases, see Supplementary material, Table S3, for statistical summaries and Supplementary material, Table S4, for post-hoc results). Confidence scores were significantly higher for vocalisations recorded at 10 and 25 metres compared to those at 50 and 75 metres, which in turn also had significantly higher confidence scores than those recorded at ≥ 100 metres (Figure 2a). For example, the mean confidence score for vocalisations within 75 metres (where detection probability was consistently high, Figure 1a) was 0.752 (N = 465 vocalisations) while scores for vocalisations recorded at ≥ 100 metres averaged 0.402 (176 vocalisations). These results suggest the possibility of modelling the relationship between confidence scores and bird distance to the recorder, which may have implications for

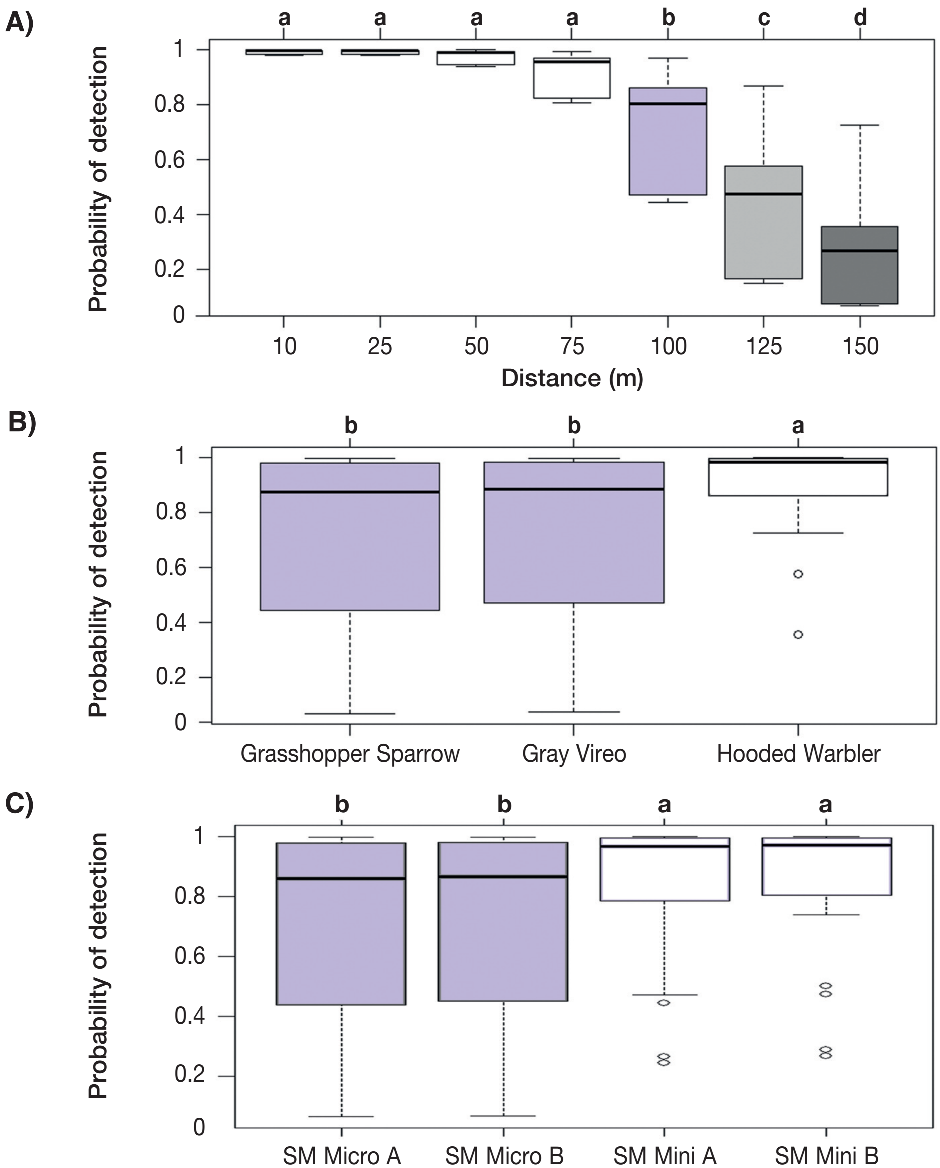

FIG. 1.—Boxplots showing the probability of BirdNET-Analyzer to correctly identify a bird vocalisation as a function of (A) bird distance to the recorder, (B) bird species, and (C) recorder. Graph calculations are based on the logistic regression included in Supplementary material, Table S1. Different colours and letters indicate significant differences in detection success after Tukey test (Supplementary Table S2).

[Diagramas de cajas mostrando la probabilidad de BirdNET-Analyzer en identificar correctamente una vocalización de un ave en función de (A) la distancia del ave al grabador, (B) especie de ave, y (C) grabador. Los cálculos de las gráficas se basan en la regresión logística incluida en la Tabla S1 del Material suplementario. Los colores y letras diferentes muestran diferencias significativas en la probabilidad de identificación según el test de Tukey (Tabla Suplementaria S2).]

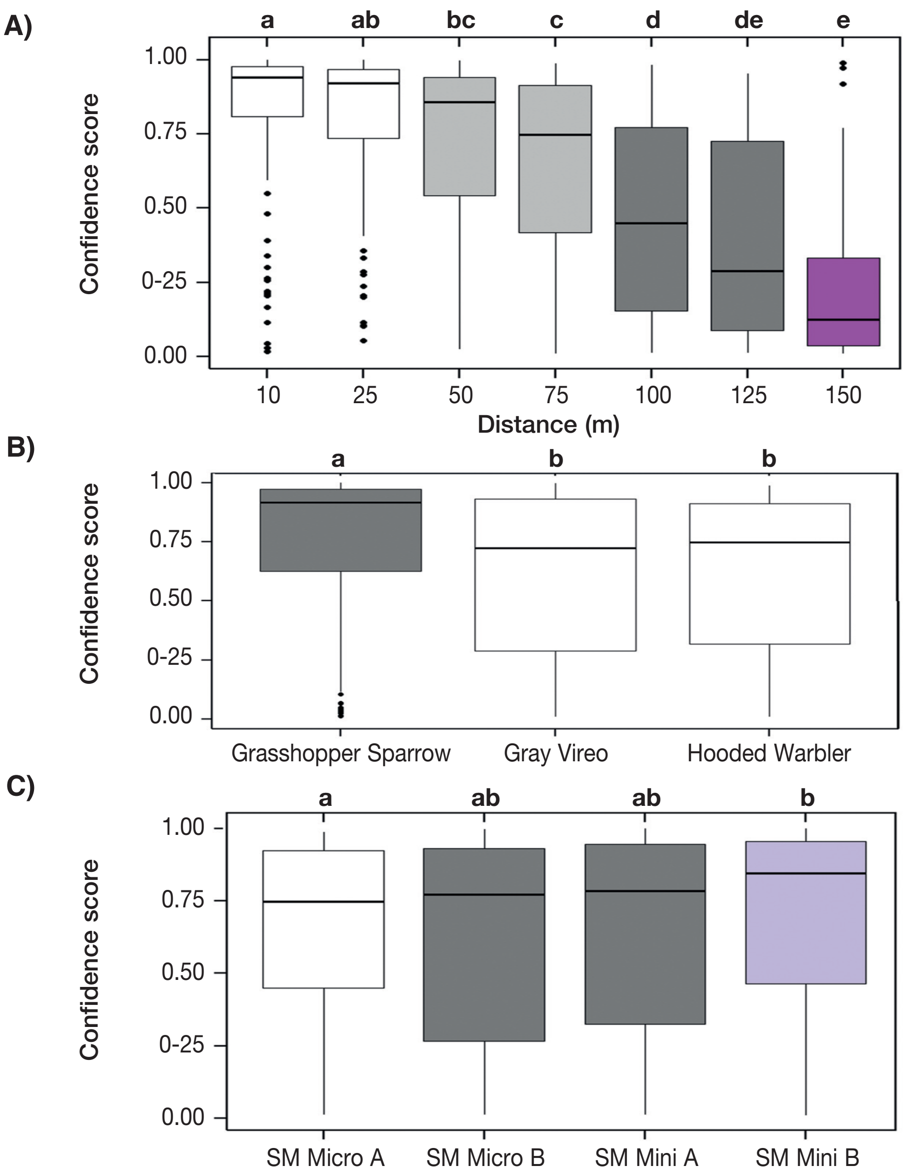

FIG. 2.—Boxplots showing the BirdNET-Analyzer confidence scores for detected bird vocalisations as a function of (A) bird distance to the recorder, (B) bird species, and (C) recorder. Graph calculations are based on the logistic regression included in Supplementary material, Table S3. Different colours and letters indicate significant differences in detection success after Tukey test (Supplementary Table S4).

[Diagramas de cajas mostrando las puntuaciones de confianza de BirdNET-Analyzer para las vocalizaciones de aves detectadas en función de (A) la distancia del ave al grabador, (B) especie de ave, y (C) grabador. Los cálculos de las gráficas se basan en el modelo de regresión beta incluido en la Tabla S3 del Material suplementario. Los colores y letras diferentes muestran diferencias significativas en la probabilidad de identificación según el test de Tukey (Tabla Suplementaria S4).]

model detection accuracies and estimating bird richness or bird density around recorders (Darras *et al*., 2016; Pérez-Granados *et al*., 2021), among other applications. This approach aligns with prior research that delimited a fixed radius around recorders modelling the relationship between recogniser confidence scores and the energy of bird vocalisations (Knight e*t al*., 2020) or that approximated the bird distance to the recorder by modelling the relationship between the energy of bird vocalisations and known bird distance to the recorder (e.g. Sebastián-González *et al*., 2018; Hedley *et al*., 2021).

A common practice when using BirdNET-Analyzer on large acoustic datasets is setting an optimal confidence score threshold to retain only high reliable predictions (Wood *et al*., 2023; Singer *et al*., 2024; Amorós-Ausina *et al*., 2024). The findings of this research suggest that setting a high confidence score threshold reduces the recorder's effective detection radius, as vocalisations recorded from greater distances have lower confidence scores. This trade-off between BirdNET-Analyzer's precision and the range of detection may complicate bird richness and bird density estimation if the detection radius is overly limited. Nonetheless, it is important to note that the recordings used here were collected under favourable recording conditions (i.e. low background noise and loudspeaker facing the recorders, Pérez-Granados *et al*., 2019). Therefore, although the same pattern may be expected to occur in further research, the specific values of the confidence scores at variable distances presented here should be used as illustrative and should not be generalised to other research settings.

BirdNET-Analyzer's confidence scores varied significantly among target species and recorders. Predictions for the Grasshopper Sparrow had higher confidence scores compared to the other two species (Figure 2b), while recordings made with the Song Meter Mini B device yielded higher confidence scores than those made with the Song Meter Micro A recorder (Figure 2c, see Supplementary material, Table S2). Interestingly, although Grasshopper Sparrow had significantly higher confidence scores (Figure 2b), the Hooded Warbler had the same probability of detection by BirdNET-Analyzer as the Grasshopper Sparrow (Figure 1b). These results suggest that BirdNET-Analyzer detection probability and model confidence are not always aligned, differences existing among species in their probability of being detected and the model's certainty of correct identification. While this experiment focused on a limited number of species and recorders, it highlights that the estimation of optimal confidence score thresholds are species- and context-dependent and cannot be directly extrapolated across species or even between studies working with the same species and recorder type, since there were also differences, although not significant, between different devices of the same recorder type (Figure 2c). Indeed, the influence of recorders and the relationship with distance also varied between species. For example, the Hooded Warbler predictions had higher confidence scores, especially for those recordings made with the Song Meter Micro, when compared to the other species (see Supplementary material, Figure S1). Moreover, environmental background, such as background noise or vocalisations from other species, may also affect BirdNET-Analyzer's performance.

In conclusion, BirdNET-Analyzer demonstrated superior performance compared to BirdNET-Api, detecting more bird vocalisations and maintaining consistently high detection probabilities over greater distances. Given these findings, researchers and managers are strongly discouraged from using the BirdNET-Api platform, as it risks producing misleading results. BirdNET-Analyzer's confidence scores vary significantly with distance, target species, and recorders, offering valuable insights for future passive acoustic bird surveys. The significant decrease in confidence scores with increasing distance between the bird and the recorder can help to identify a detection radius around the recorder, which may, for example, contribute to improving bird density estimations or allow better comparisons among studies. Further research using the same recorders and protocols should evaluate whether optimal confidence score thresholds vary among sites, which could have significant implications for large-scale bird monitoring programmes using BirdNET, which may require setting confidence score thresholds.

ACKNOWLEDGEMENTS.—I would like to acknowledge Wildlife Acoustics and Weboryx for providing me with the devices. I thank Ben McEwen and Robert Manzano-Rubio for discussing the results of the original article and encouraging me to make this reply. I am also grateful to the Associate Editor and three anonymous reviewers whose comments contributed to improving the manuscript. This research benefited from project 2021-SGR 00302 funded by Departament de Recerca i Universitats de la Generalitat de Catalunya.

REFERENCES

Amorós-Ausina, D., Schuchmann, K.L., Marques, M.I. & Pérez-Granados, C. (2024). Living together, singing together: Revealing similar patterns of vocal activity in two tropical songbirds applying BirdNET. *Sensors*, 24: 5780.

Bota, G., Manzano-Rubio, R., Catalán, L., Gómez-Catasús, J. & Pérez-Granados, C. (2023). Hearing to the unseen: AudioMoth and BirdNET as a cheap and easy method for monitoring cryptic bird species. *Sensors*, 23: 7176.

Darras, K., Pütz, P., Rembold, K. & Tscharntke, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, 201: 29-37.

Darras, K.F., Deppe, F., Fabian, Y., Kartono, A.P., Angulo, A., Kolbrek, B., Mulyani, Y.A. & Prawiradilaga, D.M. (2020). High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ*, 8: e9955.

Fox, J. & Weisberg, S. (2019). An R companion to applied regression (3rd ed.). Sage.

Funosas, D., Barbaro, L., Schillé, L., Elger, A., Castagneyrol, B. & Cauchoix, M. (2024). Assessing the potential of BirdNET to infer European bird communities from large-scale ecoacoustic data. *Ecological Indicators,* 164: 112146.

Hedley, R.W., Wilson, S.J., Yip, D.A., Li, K. & Bayne, E.M. (2021). Distance truncation via sound level for bioacoustic surveys in patchy habitat. *Bioacoustics*, 30: 303-323.

Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50: 346-363.

Jackman, S. (2024). *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory*. University of Sydney, Sydney, Australia. R package version 1.5.9.

Kahl, S., Wood, C.M., Eibl, M. & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236.

Knight, E.C., Sòlymos, P., Scott, C. & Bayne, E.M. (2020). Validation prediction: a flexible protocol to increase efficiency of automated acoustic processing for wildlife research. *Ecological Applications*, 30: e02140.

Lahoz-Monfort, J.J. & Magrath, M.J. (2021). A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience*, 71: 1038-1062.

Lenth, R. (2025). emmeans: Estimated Marginal Means, aka Least-Squares Means (R package version 1.10.7).

Lüdecke, D. (2021). performance: An R package for assessing regression models. Journal of Open Source Software, 6: 3139.

Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E. & Pérez-Granados, C. (2022). Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. *Ecological Informatics*, 72*:* 101910.

Pérez-Granados, C. (2023a). BirdNET: applications, performance, pitfalls and future opportunities. *Ibis*, 165: 1068-1075.

Pérez-Granados, C. (2023b). A first assessment of BirdNET performance at varying distances: A playback experiment. *Ardeola*, 70: 257-269.
Pérez-Granados, C., Barrero, A., Traba, J., Bustillo-de La Rosa, D., Reverter, M. & Gómez-Catasús, J. (2021). Assessment of cue counting for estimating bird density using passive acoustic monitoring: recommendations for estimating a reliable cue rate. *Avian Conservation and Ecology,* 16: 11.
Pérez-Granados, C., Bota, G., Giralt, D., Albarracín, J. & Traba, J. (2019). Cost-effectiveness assessment of five audio recording systems for wildlife monitoring: differences between recording distances and singing direction. *Ardeola*, 66: 311-325.
R Core Team (2019). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
Sebastián-González, E., Camp, R.J., Tanimoto, A.M., de Oliveira, P.M., Lima, B.B., Marques, T.A. & Hart, P.J. (2018). Density estimation of sound-producing terrestrial animals using single automatic acoustic recorders and distance sampling. *Avian Conservation and Ecology,* 13: 7.
Singer, D., Hagge, J., Kamp, J., Hondong, H. & Schuldt, A. (2024). Aggregated time-series features boost species-specific differentiation of true and false positives in passive acoustic monitoring of bird assemblages. *Remote Sensing in Ecology and Conservation*, 10: 517-530.
Stowell, D. (2022). Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 10: e13152.
Wood, C.M. & Kahl, S. (2024). Guidelines for appropriate use of BirdNET scores and other detector outputs. *Journal of Ornithology*, 165: 777-782.
Wood, C.M., Kahl, S., Chaon, P., Peery, M.Z. & Klinck, H. (2021). Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. *Methods in Ecology and Evolution,* 12: 885-896.
Wood, C.M., Kahl, S., Rahaman, A. & Klinck, H. (2022). The machine learning-powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *Plos Biology*, 20: e3001670.
Zeileis, A., Cribari-Neto, F., Gruen, B. & Kosmidis, I. (2023). *betareg: Beta Regression in R*. R package version 3.1-4.

Supplementary Electronic Material

Additional supporting information may be found in the online version of this article. See volume 72(2) on www.ardeola.org

[Información adicional sobre este artículo en su versión en línea en www.ardeola.org, volumen 72(2).]

Table S1. Summary table of the results of the logistic regression testing the relationship between BirdNET-Analyzer probability of correctly identifying a broadcast vocalisation and Distance (seven levels), Recorder ID (four levels) and Species (three levels).
[Tabla resumen de los resultados de la regresión logística que prueba la relación entre la probabilidad de BirdNET-Analyzer de identificar correctamente una vocalización emitida y la distancia (siete niveles), la identificación de la grabadora (cuatro niveles) y la especie (tres niveles).]

Table S2. Summary table of the results of Tukey post-hoc tests comparing the effect of Distance (seven levels), Recorder ID (four levels), and Species (three levels) on BirdNET-Analyzer's probability of correctly identifying a broadcast vocalisation.
[Tabla resumen de los resultados de las pruebas post-hoc *de Tukey que comparan el efecto de la distancia (siete niveles), la identificación del registrador (cuatro niveles) y la especie (tres niveles) en la probabilidad de BirdNET-Analyzer de identificar correctamente una vocalización emitida.]*

Table S3. Summary table of the results of the beta regression model testing the relationship between BirdNET-Analyzer confidence scores for predicted vocalisations and Distance (seven levels), Recorder ID (four levels) and Species (three levels).
[Tabla resumen de los resultados del modelo de regresión beta que evalúa la relación entre las puntuaciones de confianza de BirdNET-

Analyzer para las vocalizaciones predichas y la distancia (siete niveles), la identificación del registrador (cuatro niveles) y la especie (tres niveles).]

Table S4. Summary table of the results of Tukey post-hoc tests comparing the impact of Distance (seven levels), Recorder ID (four levels), and Species (three levels) on BirdNET-Analyzer's confidence scores. Graphical representations of these comparisons can be found in Figure 2. *[Tabla resumen de los resultados de las pruebas* post-hoc *de Tukey que comparan el impacto de la distancia (siete niveles), la identificación del registrador (cuatro niveles) y la especie (tres niveles) en los índices de confianza de BirdNET-Analyzer.]*

Figure S1. Boxplots showing the probability of BirdNET-Analyzer to correctly identify a bird vocalisation as a function of bird distance to the recorder, bird species, and recorder. *[Diagramas de cajas mostrando la probabilidad de BirdNET-Analyzer en identificar correctamente una vocalización de un ave en función de la distancia del ave al grabador, especie de ave, y grabador.]*

Received: January 01, 2025
Major revision: February 12, 2025
Accepted: March 09, 2025

Editor: Davide Scridel