ACOUSTIC BIRD ABUNDANCE ESTIMATION USING MORPHOLOGICAL FEATURES OF THE SPECTROGRAM

Technical Report

Yuneva E. Leon Rojas[1]*, Alejandro C. Frery*[2]*, Juan G. Colonna*[1]

[1]Instituto de Computação, Universidade Federal do Amazonas (UFAM), Manaus, Brazil
[2]School of Mathematics and Statistics, Victoria University of Wellington (VUW), New Zealand
{yunevda, juancolonna}@icomp.ufam.edu.br alejandro.frery@vuw.ac.nz

ABSTRACT

A system for estimating bird population abundance from passive acoustic recordings in zoo aviaries with multiple co-resident species is presented. The system addresses Task 6 of the BioDCASE Challenge 2026, estimating the number of individuals of three target species: Greater flamingo (Phoenicopterus roseus), Red-billed quelea (Quelea quelea), and Hadada ibis (Bostrychia hagedash). The main challenge is the acoustic saturation of the Greater flamingo: when many individuals vocalize synchronously, the number of detections stops scaling with the population. A new acoustic feature based on mathematical morphology applied to the spectrogram is proposed: the band-filtered stochastic ratio, which measures the proportion of diffuse energy in the flamingo vocalization band (400–1500 Hz). This feature directly captures acoustic saturation without relying on the species detector. The combination of the stochastic ratio with the flock-corrected detection rate (fCWR) from the official baseline reaches a correlation of $r = 0.9978$ with the real population in the four development aviaries. Using leave-one-out (LOO) validation and linear regression, the system obtains an MAE of 2.25 in the development set, which represents an 80% improvement over the official baseline ($MAE = 11.50$). For Red-billed quelea and Hadada ibis, the baseline models remain unchanged, reaching an MAE of 0 in the development set.

***Index Terms*—** Abundance estimation, mathematical morphology, Stochastic ratio, Passive acoustic monitoring

1. INTRODUCTION

The estimation of bird abundance from passive acoustic recordings represents a fundamental challenge in biodiversity monitoring. This task consists of inferring the number of individuals present in a complex acoustic scene. The problem becomes especially difficult when multiple individuals of the same species vocalize simultaneously, generating saturated acoustic signals that conventional detectors cannot reliably quantify.

This work addresses Task 6 of the BioDCASE Challenge 2026, website [1], which proposes the estimation of population abundance of three target species in European zoo aviaries: Greater flamingo (Phoenicopterus roseus), Red-billed quelea (Quelea quelea), and Hadada ibis (Bostrychia hagedash). The development set contains 140,899 3 second audio clips, distributed across 6 aviaries with known populations ranging from 4 to 161 individuals. The evaluation set contains 229,467 clips distributed across 6 main aviaries with unknown populations [2].

The species detection-based approach, such as ARIA [3], computes detection rates that work adequately when the population is small. However, when the population is large, as in the case of the Greater flamingo with more than 100 individuals, the chorus of simultaneous vocalizations saturates the detector, and the detection rate stops scaling linearly with the population. This acoustic saturation phenomenon is the main challenge of the system.

Mathematical morphology in image processing studies the geometric structures of objects and their spatial relationships. It represents images as sets of pixels arranged in two-dimensional structures and applies mathematical operations to enhance relevant shapes, facilitating their detection and recognition [4]. Applied to the spectrogram, this technique allows the decomposition of the acoustic signal into components with different morphologies: horizontal structures (sustained vocalizations), vertical structures (transient attacks), and a diffuse residual component, corresponding to the stochastic component, which reflects background noise and the overlap of multiple simultaneous vocalizations [5].

Inspired by the application of mathematical morphology [5], we propose the stochastic band ratio (estoc_banda) as a new acoustic feature that measures the proportion of diffuse energy in the flamingo vocalization band (400–1500 Hz). Combined with the flock-corrected detection rate (fCWR) from the official baseline [6], this feature enables population estimation with a correlation of $r = 0.9978$ and an MAE of 2.25 in the development set, representing an 80% improvement over the baseline (MAE = 11.50). For Red-billed quelea and Hadada ibis, the baseline models remain unchanged, reaching an MAE of 0 in the development set.

2. SYSTEM DESCRIPTION

With the purpose of estimating the abundance of three target bird species (Greater flamingo, Red-billed quelea, and Hadada ibis) from passive acoustic recordings in multispecies zoo aviaries, a system based on the stochastic band ratio is proposed, consisting of two stages:

2.1. Development stage

In the first part of the development stage (Figure 1), the system seeks to extract the features fCWR for Greater flamingo (dev_aviary_2, dev_aviary_4, dev_aviary_5, and dev_aviary_6), confidence-weighted rate (CWR) for Red-billed quelea (dev_aviary_1 and dev_aviary_3), and bout_rate_per_hour for

Hadada ibis (dev_aviary_2 and dev_aviary_4), which are necessary for population estimation.

For this purpose, the ARIA detector [3] is applied to the audio recordings from the six development aviaries (dev_aviary_1 to dev_aviary_6) using the ZooCustom_v1 model with 87 species, generating CSV files with the detected species, confidence levels, and timestamps per clip. These files are processed with feature_builder.py from the official baseline [6], together with aviary_config.json and ground_truth.csv, to extract these three acoustic features per aviary and target species, which are stored in stage2_features.csv.

For the Red-billed quelea and Hadada ibis species, the ARIA baseline model remains without modifications. However, for Greater flamingo, the baseline presents a $\mathrm{MAE} = 23.00$, indicating that the detection rate alone does not scale adequately with the population size. For this reason, as part of our proposal, the fCWR is combined with the stochastic band ratio (estoc_banda).

The calculation of estoc_banda, Figure 1, begins by applying a band-pass filter to remove the acoustic energy outside the flamingo vocalization range (400–1500 Hz) in each recording. Next, the filtered spectrogram of each clip is decomposed into three morphological components [5]: sustain (horizontal structures), attack (vertical structures), and stochastic (diffuse residue). This separation uses the scikit–image library [1], specifically the functions skimage.morphology.opening() and skimage.morphology.rectangle().

For the sustain component, a horizontal rectangular structuring element of 1 row $\times$ 20 columns is used, which allows the capture of structures extended in time but narrow in frequency, such as a sustained vocalization (long horizontal line). The 20 columns allow the detection of structures with a minimum duration of approximately 0.43 seconds, favoring the identification of real vocalizations over noise. For the attack component, a vertical rectangular structuring element of 10 rows $\times$ 1 column is used, which allows the capture of structures extended in frequency but short in time, such as the onset of a vocalization (vertical line).

The structured component (S_{struct}) is obtained as the maximum between the sustain and attack components, representing the defined structures in the spectrogram, whether horizontal or vertical. Therefore, the stochastic component per clip (S_{estoc}) corresponds to the residual obtained by subtracting S_{struct} from the filtered spectrogram, representing the diffuse energy associated with background noise and the overlap of multiple simultaneous vocalizations.

From these components, the stochastic band ratio is calculated, whose value ranges from 0 to 1 and represents the proportion of diffuse energy with respect to the total energy within the flamingo vocalization frequency band (400–1500 Hz), calculated as:

$$\mathrm{estoc\_banda} = \frac{\sum \left[\max(S_{\mathrm{estoc}}, 0)\right]^2}{\sum S_{\mathrm{band}}^2} \tag{1}$$

Where S_{band}^2 represents the total energy of those spectral coefficients. The operator $\max(S_{\mathrm{estoc}}, 0)$ retains only the positive values of the stochastic component, discarding the negative values generated during the subtraction when the structured component locally overestimates the energy. Therefore, only the positive diffuse energy contributes to the numerator. This normalized ratio allows the comparison of clips with different levels of overall acoustic activity. As the number of flamingos vocalizing simultaneously increases, a greater fraction of energy concentrates in the diffuse component.

[1] https://scikit-image.org/

The stochastic band ratio is calculated for each .WAV clip of the aviary and subsequently averaged over the N available clips, obtaining a single representative value per aviary. The resulting values for the four flamingo aviaries from the development set are stored in estoc_banda_flamingo_dev.json.

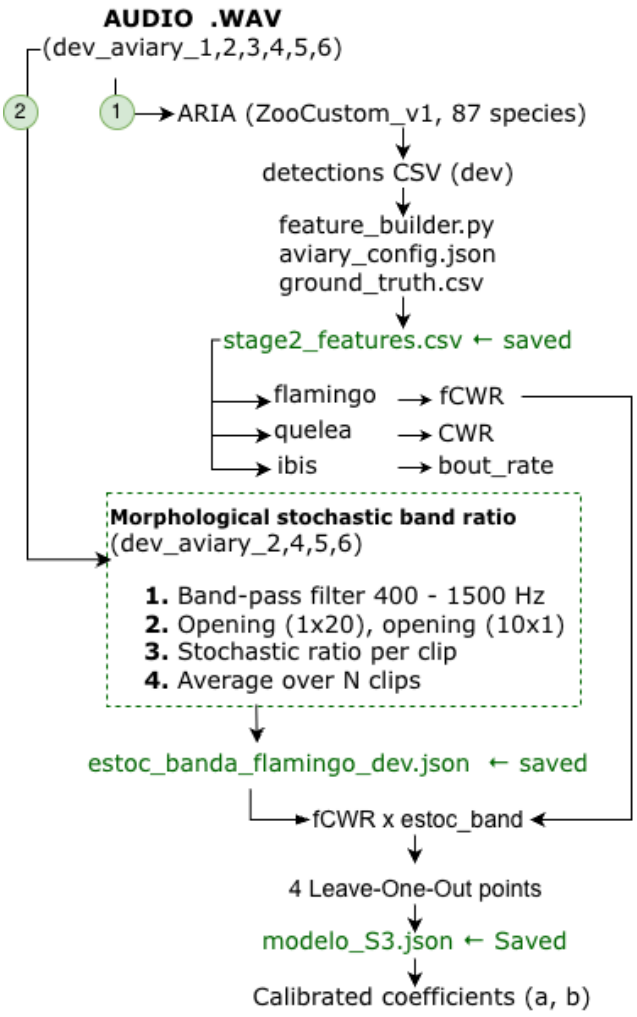

Figure 1: System calibration pipeline. ARIA detections and the morphological stochastic ratio are combined to calibrate the S3 linear model on the development aviaries.

Finally, once the $estoc\_banda$ values are obtained for each aviary, the combined feature is calculated as the product $fCWR \times \mathrm{estoc\_banda}$ for the four flamingo aviaries from the development set. Using these four points, a linear model is calibrated through least squares regression using numpypolyfit, obtaining the regression coefficients $a = 2.0342$ (slope) and $b = -21.5946$ (intercept), stored in modelo_S3json:

$$\hat{N} = 2.0342 \times (fCWR \times estoc\_banda) - 21.5946 \tag{2}$$

The model was validated using leave-one-out cross-validation (LOO) over the 4 flamingo development aviaries, obtaining an MAE of 4.50 for Greater flamingo.

2.2. Evaluation stage

In the evaluation stage (Figure 2), the ARIA detector was used on the aviaries eval_aviary_1, 2, 4, 6, 7, and 8, incorporating a filter of 67 species (eval_species_whitelist.txt) based on the eval_species_list.csv file from the challenge. This filter removes from the model the species that are not present in the evaluation aviaries, including other flamingo species (Phoenicopterus chilensis and Phoenicopterus ruber), which could generate false detections of the target species.

Subsequently, the detections were processed using a modified version of feature_builder.py, adapted for the evaluation aviaries, using the aviary_config_eval.json file generated from eval_recording_info.csv and ground_truth.csv with $count = 0$ (unknown population). The resulting features were stored in stage2_features_eval.csv.

For all audio clips corresponding to the flamingo aviaries (eval_aviary_4 and 8), the estoc_banda was calculated, and the results were stored in the file estoc_banda_eval_flamingo.json. Under the hypothesis that individual vocal behavior remains consistent between the development and evaluation stages, the combined feature ($fCWR \times \text{estoc\_banda}$) was calculated for each evaluation aviary. For this purpose, the fCWR values obtained from stage2_features_eval.csv and the estoc_banda values stored in estoc_banda_eval_flamingo.json were used.

This combined feature was applied to three prediction systems for Greater flamingo: S3, which uses a linear model, Equation 2, with the coefficients a and b stored in modelo_S3.json, previously calibrated during the development stage; S2, which corresponds to a calibrated power–law variant using the same four development points:

$$\hat{N}_{S2} = 0.6231 \times (\text{fCWR}_{\text{eval}} \times \text{estoc\_banda}_{\text{eval}})^{1.2364} \quad (3)$$

and S1, which reproduces the baseline model through the proportionality coefficient ($\text{coeff}_{\text{flamingo}}$), calculated from stage2_features.csv and ground_truth.csv from the development stage, without incorporating the stochastic component.

To predict the number of quelea individuals ($\hat{N}_{\text{quelea}}$) in the eval_aviary_1 and eval_aviary_2 aviaries, the following equation is used:

$$\hat{N}_{\text{quelea}} = \frac{\text{CWR}_{\text{eval}}}{\text{coeff}_{\text{quelea}}} \quad (4)$$

where, CWR_{eval} is extracted from stage2_features_eval.csv, calculated by feature_builder.py from the ARIA detections corresponding to each aviary. On the other hand, $\text{coeff}_{\text{quelea}}$ corresponds to a coefficient obtained during the development stage using the dev_aviary_1 and dev_aviary_3 aviaries. This coefficient is defined as the average ratio between the CWR_{dev} values (stored in stage2_features.csv) and the real population N_{dev} (recorded in ground_truth.csv):

$$\text{coeff}_{\text{quelea}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\text{CWR}_{\text{dev},i}}{N_{\text{dev},i}} = 4.35\ \text{CWR/individuo} \quad (5)$$

For Hadada ibis, the same approach is applied analogously:

$$\hat{N}_{\text{ibis}} = \frac{\text{bout\_rate}_{\text{eval}}}{\text{coeff}_{\text{ibis}}} \quad (6)$$

Using bout_rate_per_hour instead of CWR, with $\text{coeff}_{\text{ibis}} = 3.35$ bouts/individual/hour. This coefficient was calibrated using the dev_aviary_2 and dev_aviary_4 aviaries, and applied to the evaluation aviaries eval_aviary_6 and eval_aviary_7.

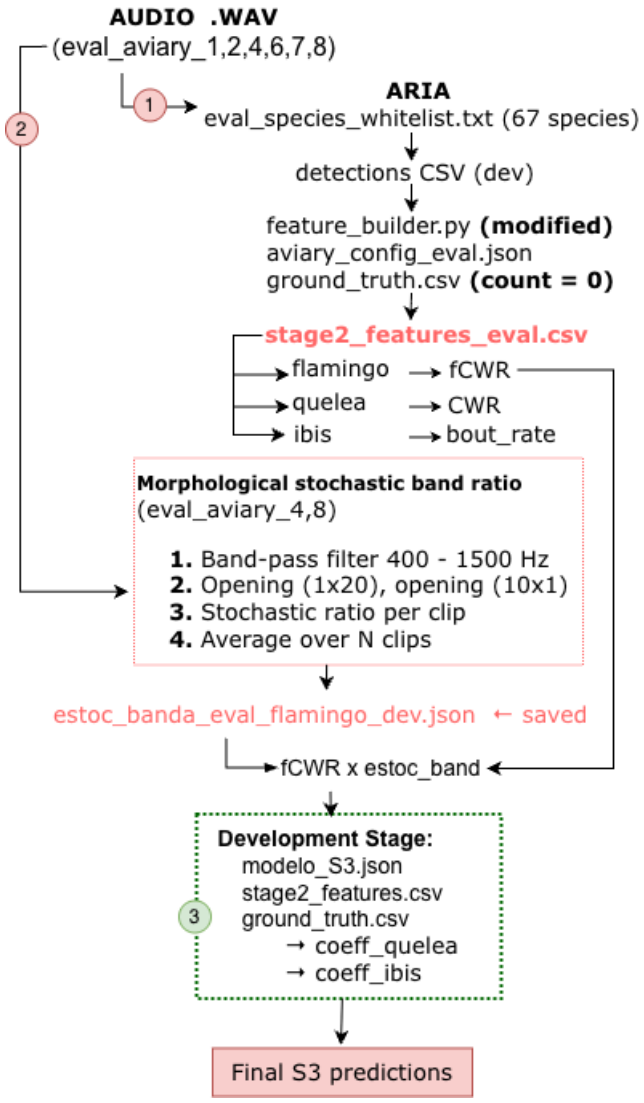

Figure 2: Prediction pipeline. The coefficients calibrated in development are applied to the new acoustic features from the evaluation aviaries to generate the final predictions.

3. RESULTS

3.1. Stochastic feature per aviary (development)

The estoc_banda calculated for all clips from each flamingo aviary during the development stage is presented in Table 1 together with the known real population. These values were used to calculate the combined feature $\text{fCWR} \times \text{estoc\_banda}$, used as the basis of the S3 model.

Once the stochastic ratio per aviary was calculated, the Pearson correlation between each feature and the real population was evaluated. As shown in Table 2, the combined feature $\text{fCWR} \times \text{estoc\_banda}$ reached $r = 0.9978$, outperforming the individual features and justifying its use as a population predictor.

Table 1: Stochastic feature and combined feature values for flamingo aviaries during the development stage.

Aviary	Real Population	estoc_banda	fCWR	fCWR × estoc_banda
dev_aviary_6	520	0.020874	1837.90	38.36
dev_aviary_5	520	0.020642	1659.14	34.24
dev_aviary_2	1070	0.018754	3360.64	63.03
dev_aviary_4	1610	0.028120	3189.92	89.70

Table 2: Pearson correlation between acoustic features and real population for Greater flamingo development aviaries.

Feature	Pearson correlation (r)
fCWR only	+0.8675
estoc_banda only	+0.7311
fCWR × estoc_banda	+0.9978

3.2. LOO validation strategy

With only 4 calibration points available (dev_aviary_2, 4, 5, and 6), standard train/test validation is not applicable. LOO validation was used, where in each iteration one aviary is left out for prediction and the model is calibrated using the remaining 3 aviaries (Table 3). This procedure ensures that each prediction is performed with a model that has not seen the aviary being predicted, simulating the real evaluation scenario.

Table 3: Leave-One-Out (LOO) validation strategy using development aviaries.

Iteration	Calibration	Prediction
1	dev_aviary_5, 2, 4	dev_aviary_6
2	dev_aviary_6, 2, 4	dev_aviary_5
3	dev_aviary_6, 5, 4	dev_aviary_2
4	dev_aviary_6, 5, 2	dev_aviary_4

3.3. LOO Results: Linear model (S3)

Table 4 shows the LOO validation results for the S3 on the 4 flamingo aviaries from the development stage. In each iteration, the model is calibrated using 3 aviaries and predicts the fourth one. The S3 model obtains an MAE of 4.50 for Greater flamingo, representing an 80% improvement over the baseline ($MAE = 23.00$).

Table 4: LOO validation results for S3 linear model on the Greater flamingo development set.

Aviary	Real Population	S3 Prediction	Error
dev_aviary_6	525	520	-2
dev_aviary_5	525	529	+4
dev_aviary_2	1070	1063	-7
dev_aviary_4	1610	1607	-3
MAE (flamingo)			**4.50**

3.4. Comparison of models in the development stage

Table 5 compares the three proposed systems (S1, S2, and S3) using the MAE in the development stage. The differences between the systems are due only to the model applied for Greater flamingo. For the quelea and ibis, the baseline model is maintained, achieving an MAE of 0.

Table 5: Comparison of submissions on the development set.

Submission	Flamingo Model	Flamingo MAE	Total MAE
S1	Baseline fCWR	23.00	11.50
S2	Power fCWR	6.25	3.12
S3	Linear fCWR	4.50	2.25

As shown in Table 5, the S2 and S3 systems outperform the baseline. However, the S3 linear model provides a better fit with only four calibration points.

3.5. Evaluation predictions

The final predictions for the 6 aviaries of the main ranking, generated by the S1, S2, and S3 systems, are presented in Table 6. The differences between the systems are observed only in the Greater flamingo aviaries (eval_aviary_4 and eval_aviary_8).

Table 6: Final predictions for evaluation aviaries.

Aviary	Species	S1	S2	S3
eval_aviary_1	Red-billed quelea	158	158	158
eval_aviary_2	Red-billed quelea	68	68	68
eval_aviary_4	Greater flamingo	422	418	418
eval_aviary_6	Hadada ibis	13	13	13
eval_aviary_7	Hadada ibis	17	16	16
eval_aviary_8	Greater flamingo	109	156	155

For eval_aviary_8, the S2 and S3 predictions converge at approximately 155 individuals, since the combined feature (87.038) falls within the development calibration range (34.2—89.7). In contrast, for eval_aviary_4, the predictions show greater divergence (18-–42 individuals), because the combined feature (19.425) falls outside the calibration range, representing an extrapolation with higher uncertainty.

4. CONCLUSIONS

A bird abundance estimation system for zoo aviaries based on passive acoustic recordings was proposed. The main contribution is the band-filtered stochastic ratio, a new feature based on mathematical morphology that captures the diffuse energy of the chorus of simultaneous vocalizations in the 400-–1500 Hz band. Combined with the baseline fCWR, this feature achieves a correlation of $r = 0.9978$ with the real Greater flamingo population and obtains an $MAE = 2.25$ in development, representing an 80% improvement over the baseline ($MAE = 11.50$).

Among the system limitations, the following stand out: the model calibration requires development aviaries with known populations of the same species; the ibis model extrapolates outside the calibration range during evaluation; and the reduced number of calibration points (4 flamingo aviaries) limits the statistical robustness of the model.

As future work, we propose exploring the application of the stochastic feature for Red-billed quelea and Hadada ibis, expanding the calibration set with more aviaries, and investigating more robust nonlinear models for populations outside the calibration range.

5. REFERENCES

[1] https://biodcase.github.io/challenge2026/task6.

[2] E. Argın, A. Härmä, and A. Arslan-Dogan, "Biodcase 2026 bird counting: Avian population estimation from passive acoustic recordings," https://huggingface.co/datasets/Emreargin/BioDCASE2026_Bird_Counting, 2026, hugging Face dataset.

[3] E. Argın, B. Amado Pereira da Costa, A. Härmä, and A. Arslan-Dogan, "ARIA: Acoustic Recognition for Inventory in Aviaries," in *Proceedings of the IEEE World Congress on Computational Intelligence (WCCI) / International Joint Conference on Neural Networks (IJCNN)*, 2026, accepted, to appear.

[4] C. A. Santos, A. I. d. Souza, and A. von Wangenheim, "Morfologia matemática," in *Visão Computacional*, A. von Wangenheim, Ed. Florianópolis: INE-CTC-UFSC, 1998, vol. 1, pp. 50–80.

[5] G. Romero-García, I. Bloch, and C. Agón, "Mathematical morphology applied to feature extraction in music spectrograms," in *International Conference on Discrete Geometry and Mathematical Morphology*. Springer, 2024, pp. 431–442.

[6] E. Argın, A. Härmä, and A. Arslan-Dogan, "BioDCASE 2026 Bird Counting Baseline: Avian Population Estimation from Passive Acoustic Recordings," https://github.com/ml4biodiversity/biodcase-population-estimation, 2026.