Universidade Estadual de Londrina

GABRIEL LIMA MEDINA ROSA

APPLICATION OF COMPUTATIONAL METHODS TO ANALYZE VOCAL SIGNALS AND REPERTOIRES OF NEW WORLD JAYS

Londrina
2019

Universidade Estadual de Londrina

Programa de Pós-Graduação Ciências Biológicas

CAPES

GABRIEL LIMA MEDINA ROSA

APPLICATION OF COMPUTATIONAL METHODS TO ANALYZE VOCAL SIGNALS AND REPERTOIRES OF NEW WORLD JAYS

Londrina
2019

GABRIEL LIMA MEDINA ROSA

APPLICATION OF COMPUTATIONAL METHODS TO ANALYZE VOCAL SIGNALS AND REPERTOIRES OF NEW WORLD JAYS

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Doutor em Ciências Biológicas (Biodiversidade e Conservação de Habitats Fragmentados).

Orientador: Prof. Dr. Luiz dos Anjos

Londrina
2019

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

R788 Rosa, Gabriel Lima Medina.

Application of computational methods to analyze vocal signals and repertoires of New World Jays / Gabriel Lima Medina Rosa. - Londrina, 2019.
112 f. : il.

Orientador: Luiz dos Anjos.
Tese (Doutorado em Ciências Biológicas) - Universidade Estadual de Londrina, Centro de Ciências Biológicas, , 2019.
Inclui bibliografia.

1. Bioacústica - Tese. 2. Repertório Vocal - Tese. 3. Gralhas - Tese. I. Anjos, Luiz dos. II. Universidade Estadual de Londrina. Centro de Ciências Biológicas. . IV. Título.

CDU 574

Dedico este trabalho à minha amada esposa, Jerusa, e à minha amada filha, Sofia, excelentes razões para que eu seja mais do que sou.

AGRADECIMENTOS

Agradeço ao meu orientador, Luiz dos Anjos, pelos valiosos ensinamentos e parceria desde a Iniciação Científica, pela liberdade de fazer tentativas, erros e acertos, mas, acima de tudo, por me contagiar com o gosto pela Ornitologia.

À CAPES, ao CNPq, Fundação Araucária e CONFAP pelo apoio financeiro que possibilitou esse estudo.

Ao professor Marco Gamba, professor Olivier Friard e colegas do Eto Lab, por me receberem na UNITO e proporcionarem uma experiência rica em conhecimento, programação, algoritmos, boa música, queijos (italianos *vs*. franceses) e trilhas na montanha.

Ao professor Marcos R. Lima, pelo grande volume de conhecimento sobre estatística provida de significado e contribuições significativas durante diversos momentos durante o projeto, incluindo o exame de Qualificação.

Ao Dr. Carlos B. de Araújo, pelas valiosas sugestões e *insights*, com as ideias certas nas horas certas.

Ao PhD. Marcelo A. Salas, PhD. David D. Yandell e PhD. Emma Greig por contribuírem com ideias e com trechos de código fundamentais para a realização deste estudo.

A todos os amigos de laboratórios que não compartilham somente refeições no “RU”, mas a empolgação em pensar, aprender e em se superar.

Ao meu pai, Marco Antônio, por ensinar a viver de forma leve e criativa, e à minha mãe, Aline, pelo valor da vida em família que torna a vida mais rica, mas a ambos, acima de tudo, por me apresentarem o amor excede todo entendimento.

À família mais amorosa e divertida que eu poderia ter, pelas muitas risadas, amor, encorajamento e por compartilharem suas vidas.

GABRIEL LIMA MEDINA ROSA

APPLICATION OF COMPUTATIONAL METHODS TO ANALYZE VOCAL SIGNALS AND REPERTOIRES OF NEW WORLD JAYS

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Estadual de Londrina, como requisito parcial à obtenção do título de Doutor em Ciências Biológicas (Biodiversidade e Conservação de Habitats Fragmentados).

BANCA EXAMINADORA

____
Orientador: Prof. Dr. Luiz dos Anjos
Universidade Estadual de Londrina – UEL

____
Prof. Dr. Luis Felipe Toledo
Universidade Estadual de Campinas – UNICAMP

____
Prof. Dr. Fernando Camargo Jerep
Universidade Estadual de Londrina – UEL

____
Prof. Dr. Robalinho Lima
Universidade Estadual de Londrina – UEL

____
Prof. Dr. Carlos Barros de Araujo
Universidade Estadual de Londrina – UEL

Londrina, 29 de abril de 2019.

ROSA, Gabriel Lima Medina. **Aplicação de métodos computacionais para analisar sinais e repertórios vocais das Gralhas do Novo Mundo**. 2019. 112 f. Tese (Doutorado em Ciências Biológicas) - Universidade Estadual de Londrina, Londrina, 2019.

RESUMO

Poucos componentes do comportamento animal podem ser registrados tão integralmente quanto vocalizações. Produzir conhecimento a partir destas é possível por meio de inferências, descobertas de padrões e predições, mas para isso são necessárias as ferramentas computacionais adequadas. Em ambos os capítulos desta Tese, exemplifico como análises acústicas e outros métodos podem ser explorados de forma robusta. No capítulo primeiro, meu objetivo foi explorar as consequências de diferentes combinações de métodos para representar sons e de análise de agrupamento para acessar o nível de organização em agrupamento segundo suas características acústicas. Por meio de análise de agrupamento, demonstrei uma nova estratégia para descrever a estrutura de repertórios vocais complexos de forma replicável mesmo na ausência e informações prévias além das características acústicas das vocalizações analisadas. Para este fim, utilizei como exemplo os repertórios de três espécies de gralhas, *Cyanocorax chrysops*, *C. coeruleus* e *C. morio*. No segundo capítulo, demostrei como abordagens computacionais essencialmente diferentes podem ser utilizadas de forma complementar para investigar variação geográfica em vocalizações de *C. coeruleus*. Ao comparar vocalizações de diferentes localidades, verifiquei relação direta entre aumentos em diferenças entre vocalizações em função de aumentos em suas distâncias geográficas, além de serem maiores que o esperado ao acaso em comparações entre as populações a Leste e Oeste da Serra do Mar. Aprimorei esta inferência de forma empírica por meio de predições da população de origem de cada amostra, segundo suas características acústicas. Por fim, discuti possíveis implicações destes resultados para a conservação desta espécie de gralha ameaçada, e sua conexão com a Mata Atlântica no Sul do Brasil e sua história, instável desde a última glaciação.

Palavras-chave: *Cyanocorax.* Gralha. Repertório vocal. Variação fenotípica. Vocalizações.

ROSA, Gabriel Lima Medina. **Application of computational methods to analyze vocal signals and repertoires of New World Jays**. Thesis (Doctoral degree in Biological Sciences) - Universidade Estadual de Londrina, Londrina, 2019.

ABSTRACT

Few components of animal behaviour can be recorded as fully as vocalizations. Producing knowledge about it is possible through inferences, pattern discovery, and predictions, but for such, adequate computational tools are required. In both chapters of this thesis, I exemplified how acoustic analysis and other methods can be exploited robustly. In the first chapter, my objective was to explore the consequences different combinations of methods for sound representation and cluster analysis to assess how vocal repertoires cluster according to their acoustic features. By means of cluster analysis, I demonstrated a new strategy to describe the structure of complex vocal repertoires in a reproducible way, even in the absence and previous information besides the acoustic features from the analysed vocalizations. For this purpose, I used three species of jays as examples: *Cyanocorax chrysops*, *C. coeruleus* and *C. morio*. In the second chapter, I demonstrated how essentially different computational approaches can be used in a complementary way to investigate geographic variation in vocalizations of *C. coeruleus*. While comparing vocalizations from different locations, I verified the direct relationship between increases in differences between vocalization and increases in geographic distance, and also larger than expected by chance in comparisons between populations East and West of the Serra do Mar. I improved this inference empirically trough predictions of the population of origin of each sample according to their acoustic features. Finally, I discussed possible implications of these results for the conservation of this threatened Jay species, and their connection to the Atlantic Forest in Southern Brazil and its unstable history of the since the last glaciation.

Key words: *Cyanocorax.* Jay. Phenotypic variation. Vocal repertoire. Vocalizations.

LISTA DE FIGURAS

Capítulo 1

Capítulo 2

LISTA DE TABELAS

Capítulo 1

Capítulo 2

SUMÁRIO

INTRODUÇÃO GERAL

Métodos para a gravação de sons estão disponíveis desde o final do século XIX, mas após o a invenção de sonógrafos analógicos, no final da década de 1950, consolidou-se a possibilidade de medir variações de frequência em sons de forma objetiva (Marler 2004). Imediatamente, as primeiras representações de flutuações de energia de forma simultânea ao longo do tempo e do espectro de frequência, i.e. sonogramas, produzidas a partir de vocalizações de aves foram feitas pelo ornitólogo Donald Borror (Borror 1956; Borror & Reese 1953, 1956). Desde então, poucos aspectos do comportamento animal podem ser registrados tão integralmente quanto vocalizações. Nas décadas seguintes, consolidou-se o estudo da comunicação vocal em animais como uma das ferramentas mais relevantes para investigar seu comportamento (Marler 2004).

Avanços mais recentes em outras aplicações, como sequenciamento genético (Brazma et al. 2003) e imagens de radiotelescópios com milhões de corpos celestes (Edgar et al. 2010; Norris 2010) permitiu que dados de grande complexidade fossem acumulados em quantidades massivas. Atualmente, a portabilidade de equipamentos de gravação de som em alta resolução, bem como a miniaturização dos volumes necessários para o armazenamento de grandes volumes de dados têm possibilitado grandes avanços nesta área de estudo (Blumstein et al. 2011). Vide o desenvolvimento de gravadores autônomos, com capacidade de registrar meses de sons de comunidades ecológicas inteiras na ausência de observadores em campo (Blumstein et al. 2011). Neste cenário, hipóteses passíveis de teste se tornam igualmente complexas, de modo inviável sem o desenvolvimento proporcional de infraestrutura computacional, softwares e algoritmos eficientes para esta finalidade. Há quase duas décadas, Breiman (2001) defendeu que aliar análises exploratórias e modelagem estatística tradicionais a novas abordagens computacionais, como aprendizado de máquina e mineração de dados, viabilizaria a produção de novos conhecimentos, generalizações, predições e descobertas de novos padrões neste cenário. O nível de automação atingido por meio desses métodos têm possibilitado grandes avanços no monitoramento de espécies ameaçadas (Dema et al. 2018; MacLaren 2018; Wood et al. 2019), conservação de comunidades (Farina et al., 2011; Deichmann et al., 2017; Wrege et al., 2017; Sayuri et al. 2019) e comportamento (Blumstein et al., 2011; Llusia et al., 2013).

A quantidade de informações presentes em uma vocalização com duração de frações de segundos pode ser notável. Por isso, um dos primeiros passos para testes de hipótese é a compressão desse conteúdo ou a extração de um conjunto limitado de informações (Baker & Logue 2003; Stowel & Plumbey 2014; Araya-Salas et al. 2017). Uma das abordagens mais comuns é representar sons por meio de características acústicas predefinidas (‘*predefined acoustic features*’ ou PAF), compostas por dados multivariados nos quais cada valor representa um conceito separado dos demais, selecionados de acordo com a relevância em relação à hipótese testada. A facilidade de interpretação de padrões biológicos relacionados à uma dada PAF é uma de suas principais vantagens. Por exemplo, a média de frequência fundamental é frequentemente comparada à massa corporal em estudos que investigam relações alométricas entre vocalizações e morfologia (Fletcher 2004; Torres et al. 2017). Porém, uma limitação de PAF é a impossibilidade de representar padrões que variam de forma complexa e não-linear ao longo do tempo.

Por meio de séries temporais, estas variações podem ser acessadas com menos perda de informações em função do tempo, em intervalos definidos (Searby & Jouventin 2004). Por exemplo, a variação de valores de frequência fundamental em função de tempo pode ser reunida para compor uma única série temporal univariada, i.e. contorno de frequência fundamental. Séries temporais também podem representar aspectos descritos de forma mais complexa, como espectros de frequência completos, obtidos por meio do algoritmo *fft* (‘*Fast Fourier Transform*’; see Brigham & Brigham 1988), que é justamente o processo de construção de sonogramas. Como comentado anteriormente, sonogramas permitem a visualização e medição das características acústicas de forma eficiente e intuitiva, mas utiliza-los integralmente em diversos tipos de análises pode resultar em um aumento excessivo de custo computacional. Alternativas como coeficientes mel-cepstrais ou MFCC (‘*mel-frequency cepstral coefficients*’; Davis & Mermelstein 1980) permitem ao mesmo tempo simplificar a dimensionalidade da representação de um som e aproximá-la da não-linearidade da percepção biológica. Por esta razão MFCC são amplamente utilizados em *softwares* de reconhecimento de voz (Tiwari 2010) e em estudos que comparam vocalizações de conteúdo harmônico rico e modulações complexas (Cheng 2010; Mielke & Zuberbühler 2013; Elie & Theunissen 2016).

Segundo os pressupostos das análises pretendidas, é frequente a necessidade de submeter descritores de vocalizações a pré-processamento adicional. Dados multivariados compostos por PAF devem ser cautelosamente inspecionados para verificação de normalidade, multicolinearidade, detecção de *outliers*, de interações não-lineares entre variáveis, que são alguns dos pressupostos para que possam ser utilizadas para inferência em testes de relações entre variáveis, como regressões lineares (Zuur et al. 2010). Diferenças de escala entre PAF, como entre variáveis medidas em segundos e variáveis medidas em Hertz, torna necessária a padronização para que os padrões contidos nos dados não sejam mascarados por estas diferenças em análises multivariadas (Legendre & Gallagher 2001). Quando há problemas de dimensionalidade, a análise de componentes principais (PCA; Mardia et al. 1979; Venables & Ripley 2002) pode controlar a quantidade de dimensões e determina a importância de cada variável na variância total dos dados, além de reduzir o esforço computacional para processar conjuntos de dados muito grandes.

Alguns algoritmos de classificação, como redes neurais (Deecke & Janik 2005), são capazes de processar séries temporais de forma direta, mas são exceções. A maioria dos métodos disponíveis, como muitos algoritmos de análise de agrupamento, exigem pré-processamento de séries temporais. Uma das possibilidades é a computação de comparações par-a-par por meio de distorção temporal dinâmica (‘*dynamic time warping*’; DTW; Bellman & Kalaba 1959; Senin 2008). DTW tem sido utilizada frequentemente para a comparação de vocalizações de aves (Lachlan et al. 2010; Jančovič et al. 2013; Meliza et al. 2013; Tan et al. 2015), cetáceos (Kershenbaum et al. 2013; Frasier et al. 2016), canídeos (Kershenbaum et al. 2016), primatas não-humanos (Gamba et al. 2015) e primatas humanos (Fedurek & Slocombe 2011). Neste algoritmo, a diferença entre dois sinais é computada segundo a soma cumulativa das distensões e compressões necessárias para o seu máximo alinhamento. Nos casos de análises incompatíveis com comparações par-a-par, o escalonamento multidimensional não-métrico (NMDS) permite que estas sejam representadas por meio de um sistema de coordenadas com perdas mínimas de informação (Faith et al. 1987; Minchin 1987).

A produção de conhecimento por meio de dados, como das formas descritas anteriormente, pode se dar por meio de três abordagens: (1) estatística, com foco em inferências baseadas na comparação do padrão observado com o que seria esperado ao acaso, segundo as condições do experimento; (2) mineração de dados, com foco na descoberta de novos padrões em dados complexos na inexistência de hipóteses prévias; e (3) aprendizado de máquina, com foco em predições, avaliadas segundo sua acurácia para novos conjuntos de dados. Apesar destas diferenças fundamentais, muitos conceitos são compartilhados entre duas ou mais áreas, e na prática, são utilizadas de forma complementar (Breiman 2001; Bzdok et al. 2018).

Existe, de fato, grande diversidade de métodos disponíveis para o estudo da comunicação vocal em animais, desde as análises acústicas às inferências, descobertas de padrões ou predições, incluindo *software* gratuito. Vide pacotes de análise acústica ‘seewave’ (Sueur et al. 2008) e ‘warbleR’ (Araya-Salas & Smith-Vidaurre 2017), disponíveis na linguagem R (R Core Teamm 2019), e o módulo ‘librosa’ (McFee et al. 2018), em *python*

(Python Software Foundation, URL *https://www.python.org/*). Apesar disso, alguns temas nesta linha de pesquisa ainda carecem de recomendações sobre as consequências de diferentes combinações de métodos e de novas soluções computacionais que impulsionem avanços.

Rosa et al. (2016) demonstram a dificuldade de realizar comparações entre repertórios vocais complexos, como os das gralhas do Novo Mundo (*sensu* Bonaccorso et al. 2010), a partir de descrições feitas com pouco ou nenhum auxílio quantitativo. Modernizar estas descrições utilizando o mesmo conjunto de métodos computacionais é uma das contribuições mais necessárias para lidar com os vieses subjetivos. No primeiro capítulo, demonstrarei como análises acústicas e mineração de dados podem ser combinadas com esta finalidade. Foram utilizadas como exemplo três espécies de gralhas do Novo Mundo com repertórios cujos tipos de vocalizações são organizados de formas claramente distintas segundo dados disponíveis na literatura. Estas espécies são a gralha-azul (*Cyanocorax coeruleus*; Anjos & Vielliard 1993), a gralha-picaça (*Cyanocorax chrysops*; Uejima 1998) e gralha-marrom (*Cyanocorax morio*; Hardy 1969). No segundo capítulo, utilizei uma coleção mais sucinta de métodos para investigar a hipótese de variação geográfica e divergência entre duas populações da gralha-azul. Para este fim, foram utilizadas comparações par-a-par entre MFCC por meio de DTW, para garantir que os padrões de variação entre vocalizações não fossem meros artefatos de técnica ou produto de PAF selecionadas arbitrariamente. Combinei métodos de inferência estatística com aprendizado de máquina, representado pelo algoritmo *Random Forest*, para demonstrar que a combinação de abordagens representa ganhos significativos no poder de inferência e interpretabilidade dos dados.

REFERENCIAS BIBLIOGRÁFICAS

Araya-Salas, M., & Smith-Vidaurre, G. (2017). warbleR: an r package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, 8, 184-191. DOI 10.1111/2041-210X.12624

Araya-Salas, M., Smith-Vidaurre, G., & Webster, M. (2017). Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. *Bioacoustics*, 28, 57-73. DOI 10.1080/09524622.2017.1396498

Baker, M. C., & Logue, D. M. (2003). Population differentiation in a complex bird sound: A comparison of three bioacoustical analysis procedures. *Ethology*, 109, 223-242. DOI 10.1046/j.1439-0310.2003.00866.x

Bellman, R., & Kalaba, R. (1959). On adaptive control processes. *IRE Transactions on Automatic Control*, 4, 1-9. DOI 10.1109/TAC.1959.1104847

Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., … Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48, 758-767. DOI 10.1111/j.1365-2664.2011.01993.x

Borror, D.J. & C.R. Reese. 1953. The analysis of bird song by means of a vibralyser. *Wilson Bulletin*, 65, 271-276.

Borror, D.J. & C.R. Reese. 1956. Vocal gymnastics in wood thrush songs. *Ohio Journal of Science*, 56, 177-182.

Borror, D.J. 1956. Variation in Carolina Wren songs. *Auk*, 73, 211-229. DOI 10.2307/4081473

Brazma, A., Parkinson, H., Sarkans, U., Shojatalab, M., Vilo, J., Abeygunawardena, N., … Sansone, S. A. (2003). ArrayExpress - A public repository for microarray gene expression data at the EBI. *Nucleic Acids Research*, 31, 6871. DOI 10.1093/nar/gkg091

Breiman, L., & others. (2001). Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science*, 16, 199-231. DOI 10.1214/ss/1009213726

Brigham, E. O., & Brigham, E. O. (1988).*The fast Fourier transform and its applications*. Prentice Hall, Englewood Cliffs

Bzdok, D., Altman, N., & Krzywinski, M. (2018). Statistics versus machine learning. *Nature Methods*, 15, 233-234. DOI 10.1038/nmeth.4642

Daniel Meliza, C., Keen, S. C., & Rubenstein, D. R. (2013). Pitch- and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *The Journal of the Acoustical Society of America*, 134, 1407-1415. DOI 10.1121/1.4812269

Davis, S. B., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 28, 357-366. DOI 10.1109/TASSP.1980.1163420

Deecke, V. B., & Janik, V. M. (2005). Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. *The Journal of the Acoustical Society of America*, 117, 2470-2470. DOI 10.1121/1.4787454

Deichmann, J. L., Hernández-Serna, A., Delgado C., J. A., Campos-Cerqueira, M., & Aide, T. M. (2017). Soundscape analysis and acoustic monitoring document impacts of natural gas exploration on biodiversity in a tropical forest. *Ecological Indicators*, 74, 39-48. DOI 10.1016/j.ecolind.2016.11.002

Dema, T., Zhang, L., Towsey, M., Truskinger, A., Sherub, S., Kinley, … Roe, P. (2017). An investigation into acoustic analysis methods for endangered species monitoring: A case of monitoring the critically endangered white-bellied heron in Bhutan. In

Proceedings - 13th IEEE International Conference on eScience, eScience 2017. DOI 10.1109/eScience.2017.30

Edgar, R. G., Clark, M. A., Dale, K., Mitchell, D. A., Ord, S. M., Wayth, R. B., … Greenhill, L. J. (2010). Enabling a high throughput real time data pipeline for a large radio telescope array with GPUs. *Computer Physics Communications*, 181, 1707-1714. DOI 10.1016/j.cpc.2010.06.019

Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: a data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, 19, 285-315. DOI 10.1007/s10071-015-0933-6

Faith, D. P, Minchin, P. R. and Belbin, L. (1987). Compositional dissimilarity as a robust measure of ecological distance.*Vegetatio*, 69, 57-68. DOI 10.1007/BF00038687

Farina, A., Pieretti, N., & Piccioli, L. (2011). The soundscape methodology for long-term bird monitoring: A Mediterranean Europe case-study. *Ecological Informatics*, 6, 354-363. DOI 10.1016/j.ecoinf.2011.07.004

Fedurek, P., & Slocombe, K. E. (2011). Primate vocal communication: A useful tool for understanding human speech and language evolution? *Human Biology*, 83, 153-173. DOI 10.3378/027.083.0202

Fletcher, N. H. (2004). A simple frequency-scaling rule for animal communication. *The Journal of the Acoustical Society of America*, 115, 2334-2338. DOI 10.1121/1.1694997

Frasier, K. E., Elizabeth Henderson, E., Bassett, H. R., & Roch, M. A. (2016). Automated identification and clustering of subunits within delphinid vocalizations. *Marine Mammal Science*, 32, 911-930. DOI 10.1111/mms.12303

Gamba, M., Friard, O., Riondato, I., Righini, R., Colombo, C., Miaretsoa, L., … Giacoma, C. (2015). Comparative Analysis of the Vocal Repertoire of Eulemur: A Dynamic Time

Warping Approach. *International Journal of Primatology*, 36, 894-910. DOI 10.1007/s10764-015-9861-1

Jančovič, P., Köküer, M., Zakeri, M., & Russell, M. (2013). Unsupervised discovery of acoustic patterns in bird vocalisations employing DTW and clustering. *21st European Signal Processing Conference (EUSIPCO 2013)*, 1-5. IEEE.

Kershenbaum, A., Root-Gutteridge, H., Habib, B., Koler-Matznick, J., Mitchell, B., Palacios, V., & Waller, S. (2016). Disentangling canid howls across multiple species and subspecies: Structure in a complex communication channel. *Behavioural Processes*, 124, 149-157. DOI 10.1016/j.beproc.2016.01.006

Kershenbaum, A., Sayigh, L. S., & Janik, V. M. (2013). The Encoding of Individual Identity in Dolphin Signature Whistles: How Much Information Is Needed? *PLoS ONE*, 8, 1-7. DOI 10.1371/journal.pone.0077671

Lachlan, R. F., Verhagen, L., Peters, S., & ten Cate, C. (2010). Are There Species-Universal Categories in Bird Song Phonology and Syntax? A Comparative Study of Chaffinches (*Fringilla coelebs*), Zebra Finches (*Taenopygia guttata*), and Swamp Sparrows (*Melospiza georgiana*). *Journal of Comparative Psychology*, 124, 92-108. DOI 10.1037/a0016996

Legendre, P. & Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271-280. DOI 10.1007/s004420100716

Llusia, D., Gómez, M., Penna, M., & Márquez, R. (2013). Call Transmission Efficiency in Native and Invasive Anurans: Competing Hypotheses of Divergence in Acoustic Signals. *PLoS ONE*, 8, 1-16. DOI 10.1371/journal.pone.0077312

MacLaren, A. R., Crump, P. S., Royle, J. A., & Forstner, M. R. J. (2018). Observer-free experimental evaluation of habitat and distance effects on the detection of anuran and bird vocalizations. *Ecology and Evolution*, 8, 12991-13003. DOI 10.1002/ece3.4752

Mardia, K. V., J. T. Kent & J. M. Bibby (1979). *Multivariate Analysis*. Academic Press, London.

Marler, P. (2004). Science and birdsong: the good old days. In Marler, P., & Slabbekoorn, H.. (eds) *Nature's music: The science of birdsong*. Elsevier, San Diego.DOI 10.1016/B978-0-12-473070-0.X5000-2

Marler, P., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong*. *Nature's Music: The Science of Birdsong*. Elsevier, San Diego.DOI 10.1016/B978-0-12-473070-0.X5000-2

McFee, B., Raffel, C., Liang, D., Ellis, D., McVicar, M., Battenberg, E., & Nieto, O. (2018). librosa: Audio and Music Signal Analysis in Python. *Proceedings of the 14th Python in Science Conference*, (Scipy), 18-24. DOI 10.25080/majora-7b98e3ed-003

Mielke, A., & Zuberbühler, K. (2013). A method for automated individual, species and call type recognition in free-ranging animals. *Animal Behaviour*, 86, 475-482. DOI 10.1016/j.anbehav.2013.04.017

Minchin, P.R. (1987) An evaluation of relative robustness of techniques for ecological ordinations. *Vegetatio*, 69, 89-107.

Norris, R. P. (2010). Data challenges for next-generation radio telescopes. *Proceedings - 6th IEEE International Conference on e-Science Workshops, e-ScienceW 2010*, 21-24. DOI 10.1109/eScienceW.2010.13

Rosa, G. L. M., Ellis, J. M., Bonaccorso, E., & Anjos, L. dos. (2016). Friend or foe? Social system influences the allocation of signals across functional categories in the repertoires of the New World jays. *Behaviour*, *153*, 467-524. DOI 10.1163/1568539X-00003360

Sayuri, L., Sugai, M., & Llusia, D. (2018). Bioacoustic time capsules: Using acoustic monitoring to document biodiversity. *Ecological Indicators*, *99*, 149-152. DOI 10.1016/j.ecolind.2018.12.021

Searby, A., & Jouventin, P. (2004). How to measure information carried by a modulated vocal signature? *The Journal of the Acoustical Society of America*, 116, 3192-3198. DOI 10.1121/1.1775271

Senin, P. (2008). Dynamic Time Warping Algorithm Review. In *Information and Computer Science Department University of Hawaii at Manoa Honolulu*, 1-23. DOI 10.1109/IEMBS.2007.4353810

Stowell, D., & Plumbley, M. D. (2014). Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. *PeerJ*, *2*, e488. DOI 10.7717/peerj.488

Sueur J., Aubin T., Simonis C. (2008). Seewave: a free modular tool for sound analysis and synthesis. *Bioacoustics*, 18, 213-226. DOI 10.1080/09524622.2008.9753600

Tan, L. N., Alwan, A., Kossan, G., Cody, M. L., & Taylor, C. E. (2015). Dynamic time warping and sparse representation classification for birdsong phrase classification using limited training data). *The Journal of the Acoustical Society of America*, 137, 1069-1080. DOI 10.1121/1.4906168

Tiwari, V. (2010). MFCC and its applications in speaker recognition. *International Journal on Emerging Technologies*, 1, 19-22.

Torres, I. M. D., Lopez, L. C. S., & de Araújo, C. B. (2017). Allometric trends reveal distinct evolutionary trajectories for avian communication. *The Journal of the Acoustical Society of America*, 142, 1879-1887. DOI 10.1121/1.5005495

Venables, W. N. & B. D. Ripley (2002).*Modern Applied Statistics with S*. Springer-Verlag.

Wood, C. M., Popescu, V. D., Klinck, H., Keane, J. J., Gutiérrez, R. J., Sawyer, S. C., & Peery, M. Z. (2019). Detecting small changes in populations at landscape scales: a bioacoustic site-occupancy framework. *Ecological Indicators*, 98, 492-507. DOI 10.1016/j.ecolind.2018.11.018

Wrege, P. H., Rowland, E. D., Keen, S., & Shiu, Y. (2017). Acoustic monitoring for conservation in tropical forests: examples from forest elephants. *Methods in Ecology and Evolution*, 8, 1292-1301. DOI 10.1111/2041-210X.12730

CAPÍTULO 1

Mining for acoustic features: towards reproducible comparisons of vocal repertoires

Artigo a ser submetido ao periódico Methods in Ecology and Evolution – http://www.methodsinecologyandevolution.org/view/0/authorGuidelines.html

GABRIEL LIMA MEDINA ROSA (Orcid ID: 0000-0001-5332-9713)

Article type : Research Article

Editor : Dr. Jana Vamosi

Mining for acoustic features: towards reproducible comparisons of vocal repertoires

Gabriel Lima Medina Rosa[1,2]*, Marco Gamba[2], Olivier Friard[2], Jesse M. Ellis[3], Luiz dos Anjos[1]

[1]Universidade Estadual de Londrina, Departamento de Biologia Animal e Vegetal, Rod. Celso Garcia Cid, 10.011, Londrina, Brasil.

[2]University of Torino, Department of Life Sciences and Systems Biology, Via Accademia Albertina 13, 10134 Torino, Italy.

[3]Coe College, Department of Biology, 1220 First Avenue Northeast, Cedar Rapids, IA, USA

*__corresponding author__: g.rosa.bio@gmail.com

Running title: Clustering New World Jay vocalizations

Abstract

1. We can obtain insights into vocal communication of animal species by understanding whether we can recognize discrete groups of acoustically similar calls. In most available vocal repertoire descriptions this task was performed directly by observers, with biases that may limit reproducibility and comparisons across different studies. Cluster analysis is a robust alternative, but validation of results obtained in an unsupervised manner is rarely an easy task.
2. We aimed to evaluate the effect of different analytical decisions for clustering vocalizations in the study of vocal repertoires. The best alternatives were used to demonstrate how repertoire features, measured from vocalization clusters, are an effective strategy to describe and compare repertoires of different species with realistic and fully reproducible settings. For such, we used three species of New World Jays, known to have contrastingly different repertoires as examples.
3. We demonstrated potential problems of reproducibility in comparisons between multivariate sets of acoustic features processed and clustered independently, but not for features represented as time-series processed using dynamic time warping. Acoustic feature selection was the most important analytical decision. The acoustic structure described by repertoire features revealed results consistent with the literature and across different methods, even when sub-optimal clustering results were used. Results for *Cyanocorax chrysops* indicate a more diverse repertoire than those of the other two species, but with less variation inside possible vocalization types. In *C. morio* variation in the whole repertoire diversity was narrower and with absence of clear limits between vocalization types. *C. coeruleus* resembles both, with a diverse repertoire, but not so clear limits between possible vocalization types.
4. Along with animal sound collections, grows the importance of reproducibility in the analysis of large and complex data. Criteria beyond how clearly defined vocalization

clusters can be essential to compare repertoires with more than counts of vocalization types. For Bioacoustics and other application, constant effort is necessary to update method recommendations, and explore their full potential.

Key words: acoustic analysis, clustering, vocalizations, vocal repertoire, data mining, *Cyanocorax*, Neotropical jay

Introduction

For efficient communication, similar signals or messages must share essentially similar information. For this reason, if vocalizations produced by a given individual, group of individuals or species can be separated into acoustically homogeneous groups of vocalizations is vital for the answering of questions about their biology and behaviour (Mates et al. 2014). In the scientific literature, several accounts of acoustic variation in vocal repertoires are based on ‘manual’ classifications conducted directly by one or more observers. Although useful to interpret biological patterns and processes (Kershenbaum et al. 2013), such data is subject to biases due to differences between observers, which will compromise reproducibility (Rendell & Whitehead 2003). Also, manual classifications are not feasible when there are many samples, and lack intuitive solutions when variables are complex. Therefore, manual classification can be problematic for comparative studies (Burghardt *et al.* 2012; Deecke & Janik 2005).

A frequent solution to this problem is the classification of acoustic signals into groups defined a priori (Kahl *et al.* 2017; Valletta *et al.* 2017; Xie *et al.* 2018). Because reliable information to do so is rarely available for new data, unsupervised approaches, such as cluster analysis, could be feasible and reproducible solutions to define groups of vocalizations in vocal repertoires (Gamba *et al.* 2015; Riondato *et al.* 2017; Wadewitz *et al.* 2015). The term 'unsupervised' refers to the use of quantitative descriptors of the samples, such as acoustic features, to determine cluster memberships or hierarchies without known identities or classes (Henning 2016). In this data-driven approach, evaluating analytical decisions, such as feature selection, pre-processing techniques, algorithm selection, and parameter tuning is of paramount importance to avoid sub-optimal results. Also, reproducibility will be increased if researchers report in detail the steps used in the clustering methods and the reasons of their decisions.

Clustering of data is rarely straightforward and its application to Bioacoustics is no exception. Examples of studies in where such methods are thoroughly evaluated include the evaluation of universal categories and population variation in passerine songs (Lachlan *et al.* 2010, 2016), and automated methods for pairwise comparisons between sub-units of passerine songs (Große Ruse *et al.* 2016). Rosa *et al.* (2016) compared repertoire descriptions to investigate the relationship between repertoire size and sociality in New World Jays (NWJ; *sensu* Bonaccorso *et al.* 2010), and as consequence, highlighted the many caveats of comparisons between counts of vocalization types. Among those caveats is the difficulty to reconcile the descriptions, available only for 14 of the 39 NWJ species, and in all cases, vocalization types were identified directly by one or more observers (see Rosa *et al.* 2016). In fact, even when quantitative methods are used, NWJ repertoires can be challenging to analyse.

In the present study we aimed to give another step towards the reproducibility in the study of vocal repertoires, using NWJ as examples. We selected NWJ because of their vocal repertoires range from simple to complex in acoustic structure and social function (Hardy 1969; Rosa *et al.* 2016), which makes them useful models to explore the consequences of different methods for clustering. Therefore, we evaluated the consequences of analytical decisions over clustering results obtained independently for two species, C*yanocorax chrysops* (Vieillot, 1818) and *Cyanocorax morio* (Wagler, 1829). The former species has a highly diverse repertoire (Hardy 1969, Uejima 1998), and has recently been considered as one of the most diverse repertories described for NWJ (Rosa *et al*. 2016). Conversely, *C. morio* repertoire has all of its acoustic variation described as a single vocalization type (see line 5, items A and B of Figure 2 in Hardy 1969). We took advantage of the contrast between a complex and a simple repertoire to identify how decisions related to methods used for cluster analyses could affect their outcomes.

Our second aim was to demonstrate how different repertoires could be compared when using more than counts of vocalization types. We refer to the use of cluster measurements as a mean to infer repertoire characteristics, such as how variable vocalization types are and how different they are from others in the same repertoire. We compared the repertoires of *C. chrysops* and *C. morio*, and *Cyanocorax coeruleus* (Vieillot, 1818), under this approach and evaluated if results showed interspecific differences. Hardy (1969) considers *C. chrysops* as a typical ornate species, thus showing a trend towards elaborate plumage, with bright colours and ornaments (such as crests) and a rich vocal repertoire. Conversely, he described *C. coeruleus* as a typical inornate species, showing darker plumage, reduced or absent crests, and simple vocal repertoire. *Cyanocorax morio* classification is ambiguous. The description made by Anjos & Vielliard (1993) for *C. coeruleus* indicated large acoustic variation between vocalization types, which we expect to be closer to the variation found in *C. chrysops*, but also lacks clear limits between vocalization types, as is found for *C. morio*. Therefore, it is of particular interest to use these three species for an investigation of the interspecific differences and to exemplify the application of our analytical approach. Moreover, we evaluate how different acoustic analyses and clustering algorithms affect and contribute to interspecific comparisons, advancing towards higher automation and the necessary reproducibility for robust comparisons. With our results, we also provide descriptions of their acoustic structures and contribute to improve the current knowledge about NWJ vocal communication in the light of Hardy’s hypotheses.

Materials and Methods

MODEL SPECIES

We included three species of NWJ: *Cyanocorax coeruleus*, *C. chrysops*, and *C. morio*. These species share many aspects of their sociality, living in flocks of up to 10 to 15 individuals that use and defend stable territories throughout the year and reproduce cooperatively (Anjos 1991; Hale 2003; Williams 2004; Uejima *et al.* 2012). Molecular systematics placed *C. coeruleus* and *C. morio* as closely related within 'Clade A', and *C. chrysops* within 'Clade B', the two major lineages in the genus *Cyanocorax* (Bonaccorso *et al.* 2010). *Cyanocorax morio* is widespread in Central America, while *C. chrysops* and *C. caeruleus* are endemic to South America. The last two overlap their ranges in Southern Brazil, but *C. chrysops* reaches farther North into southern limits of the Amazonian and through Chaco wetlands until the base of the Andes in western Argentina. The species are group-living, reproduce cooperatively, have territorial behaviour (Lawton & Guindon 1981; Anjos 1991; Anjos & Vielliard 1993; Williams *et al.* 1994; Reinert & Bornschein 1998; Hale 2003; Williams 2004; Williams & Hale 2006; Boesing & Anjos 2014; Uejima *et al.* 2014). According to Rosa et al. (2016), *C. chrysops* and *C. coeruleus* repertoires are among the most functionally diverse among NWJ, also with more vocalizations from their known repertoires dedicated to social integration, compared to conflict resolution. *Cyanocorax coeruleus* plumage is mostly dark blue with black chest and head, with a short crest present in the front. *Cyanocorax chrysops* plumage, also blue in most of the wings and tail and black head and chest, is more complex, with creamy to yellowish white underparts, including a short crest, conspicuous yellow irises, a white tail tip and bright blue markings on the nape, above the eyes and at both sides of the jaw. *Cyanocorax morio* is mostly brown, with slightly clearer belly, no crest, but at some locations, white tail tips are conspicuous.

AUDIO RECORDING AND PROCESSING

In this study, we included the most conspicuous data currently available for the three study species. Recordings of *C. chrysops* (n = 189; 1h25m43s) were made by GLMR at a single location, Mata dos Godoy State Park (Londrina – PR, Brazil; 23.44305ºS; 51.24256ºW), with the digital recorder Roland R26 (96kHz; 24-bit) and Sennheiser MK60 microphone equipped by a parabolic reflector. Recordings of *C. coeruleus* (n = 84; 3h32m06s) were made by LA at Fazenda Santa Rita (Palmeira – PR, Brazil; 25.421780ºS, 49.949488ºW), including recordings from captive individuals, using an Uher 4400 (19 cm/s) tape recorder with a microphone and parabolic reflector. Recordings of *C. morio* (n = 50; 3h45m11s) were made by JME at Guindon Farm (Monteverde, Punta Arenas, Costa Rica; 10.297196ºS, 84.802902ºW), with digital recorder Marantz PMD690 (48kHz; 24-bit) and Sennheiser MK67 hypercardioid microphone. All recordings were stored at their original resolutions, but downsampled (or digitized) to 44.1kHz (24-bit) for the acoustic analysis. We filtered energy content below 300Hz and above 8000Hz from all recordings, because those frequencies are reportedly out of the hearing range of other corvids (Jensen & Klokker 2006). The start and end of all vocalisations of the studied species were assigned manually within its original recording by visual inspection of normalized spectrograms with the same parameters (window = Hann; ws = 1.75 ms; hop = 1.75 ms). Samples with low signal to noise ratio and overlapped with one or more signals, whatever their origin, were not included in the analyses. The available data about the repertoire of all three species indicate relatively simple syntactical structures, in which some call types can be repetitions of nearly identical acoustic units. Therefore, to ensure the adequacy of the acoustic analysis, we limited our analyses to the acoustic units. Whenever we identified calls composed of two or more identical units, easily identified because of their regularity, all but one was discarded from further analyses. This step was necessary to avoid biasing the representation of vocalizations composed by repeated elements in our data. The selected samples were cut into separate files, each containing a single acoustic signal and normalized to 0dB, each representing a vocalization or instance in the data, for clustering purposes. We conducted our analyses using 3933 calls for *C. caeruleus*, 2832 for *C. chrysops*, and 2245 for *C. morio*.

Acoustic Features

We used three different types of acoustic analysis (Table S1): predefined acoustic features (PAF), univariate time-series and multivariate time-series. We referred to each alternative description of vocalizations as feature spaces because the clustering algorithms search for clustering structure in the space defined by feature dissimilarities between signals. The first strategy, PAF, is based in the 29 acoustic variables that can be extracted from a given sound using the function 'specan' of the package '*warbleR*' v.1.1.12 (Araya-Salas & Smith-Vidaurre 2017; see Table S2 in Supp. Matt.) using the same spectrogram parameters (f = 44100; wl = 256; ovlp = 90%) and restricted to the same frequency band (300-6000Hz). To avoid the effects of differences in scale magnitudes between each PAF, such as that between variables measured in Hertz and seconds, we standardized the original set of PAF to zero mean and unit standard deviation and produced the first feature space: $fs\mathrm{PAF}_{A}$. Because multicollinearity can affect clustering results, we included two pre-processing methods: variance inflation factor (VIF; Zuur *et al.*, 2010) and principal component analysis (PCA; R package '*FactoMineR*' *v*.1.40; Husson *et al*. 2010). We kept only the variables with VIF inferior to 10 (Zuur *et al.* 2010) in $fs\mathrm{PAF}_{A}$ to generate $fs\mathrm{PAF}_{B}$. By performing PCA on $fs\mathrm{PAF}_{A}$ and on $fs\mathrm{PAF}_{B}$, we obtained $fs\mathrm{PAF}_{pcA}$ and $fs\mathrm{PAF}_{pcB}$. In both cases, the number of components kept as inputs for cluster analysis were selected using generalized cross-validation approximation (Josse & Husson 2012). We also computed Random Forest similarity matrices (Breiman 2002), and generated $fs\mathrm{PAF}_{RF.}$ This technique is based in a decision tree process for feature construction, followed by a two-class similarity learning, and is reportedly effective to improve classification performance for bird calls (Keen *et al.* 2014). Because random forest results are affected by variable ranks and not scaling, it can be performed without standardizing variables and avoids decisions about which threshold to be used for the selection of variables (i.e. VIF) or number of components to be kept (i.e., PCA). To identify the consequences of pre-processing methods while clustering PAF, we performed cluster analyses independently for the combinations of methods described above (see Table S1).

Because essential features of animal sounds may be coded as continuous variation patterns within a vocalization, we sampled such features over fixed intervals and built time-series representations. In our second strategy, we represented vocalizations using two different univariate time-series: dominant frequency contour (DFC) and fundamental frequency contour (FFC). While DFC is a regular time-series of frequency spectrum positions in which the energy peaks are located, FFC, represents the frequency values of the fundamental harmonic, calculated using the function 'FF' of the R package 'tuneR' v. 1.3.2 (Ligges 2018). Both univariate time-series were measured using the same spectrogram parameters and frequency band as PAF. Increasing the level of detail, our third strategy consisted in using multivariate time-series based in energy coefficients at different frequencies in the spectrum through time: linear frequency bins (LFB) and Mel-a non-linear equivalent, Frequency Cepstral Coefficients (MFCC; Davis & Mermelstein 1980). We measured LFB by sampling energy over fixed spectra of 5ms throughout 12 frequency bins between 501 Hz and 6000 Hz in intervals of 500 Hz. In other words, the LFB representation could be considered a representation analogous to a spectrogram with decreased resolution. The 12 cepstral coefficients of the MFCCs were calculated over fixed cepstra of 5 ms (hop = 1 ms, lift.exp = 22, pre.filt = 0.97), as a non-linear alternative to LFB.

To ensure the quality of signal comparisons made using the univariate and multivariate time-series data (Meliza *et al.* 2013; Elie *et al.* 2011; Krull *et al.* 2012; Ratanamahatana & Keogh 2004; Somervuo 2018) and adapt it for cluster analysis, we used the Dynamic Time Warping algorithm (DTW; package 'dtw' v.1.18-1; Giorgino 2009; and 'dtwclust' v5.1.0; see Sarda-Espinosa 2017 for more details). The DTW provides pairwise dissimilarities between time-series using the cumulative non-linear distortion and compression necessary for their maximal alignment. Therefore, all time-series data used for clustering, referred as *fs*DFC, *fs*FFC, *fs*LFB, and *fs*MFC, were pre-processed as DTW dissimilarities calculated between the samples obtained for each species. To preserve differences between species we preserved the original scale of DTW dissimilarities.

CLUSTERING ALGORITHMS AND PARAMETERS

We included three different clustering algorithms to explore the consequences of such analytical decisions on clustering configurations. Those algorithms are: Partitioning Around Medoids (hereafter PAM; package 'cluster' v.2.0.6; Maechler *et al.* 2017), k-means (packages 'stats'; R Core Team 2017; package 'ClusterR' v.1.1.1; Mouselimis 2018) with three different initialization strategies, and hierarchical clustering following Ward's minimum variance method (hereafter h-ward; package 'stats'; R Core Team 2017; see also Murtagh & Legendre 2014). Hierarchies obtained with h-ward, graphically displayed as a dendrogram, can be cut into k clusters to obtain and interpret the clustering structure as discrete clusters. The remainder of the algorithms find clusters using distances from all instances in the data to prototypes in the feature space. Those prototypes are iteratively moved and memberships reassign until the sum of all intra-clusters dissimilarities stabilizes at low values. With the same data and parameters, results of k-means could also vary according to the position in the feature space where the prototypes are initialized. To assess the effect of initialization strategies for k-means, we compared random initializations of k-means with k-means++ (package 'ClusterR' v.1.1.0; Mouselimis 2018; *see also* Arthur & Vassilvitskii 2007), which improves the quality of the initial prototypes, and hybrid k-means (hereafter hk-means; Kassambara 2017), which uses prototypes obtained from cuts in a h-ward dendrogram. The hk-means was implemented as a modified version of 'hkmeans', available in the package 'factoextra' v.1.0.5 (Kassambara & Mundt 2017).

All clustering algorithms included in this study require prior definition of the desired number clusters (k) to be found in the data. Because external data to use as ground-truth or estimate the most adequate number of clusters is not available, we produced alternative results with all integers in the interval $2 \leq k \leq 35$. We selected this interval because it includes all possibilities between the simplest clustering structure ($k = 2$) and the most complex scenario ($k = 35$), in which the number of vocalization types is reasonably higher than counts of vocalization types in descriptions of all three repertoires (2 in *C. morio*, 14 in *C. caeruleus* and 23 in *C. chrysops*). Therefore, each combination of the above-mentioned acoustic features, pre-processing methods and clustering algorithms, had 34 versions, each with one more cluster than the previous. Additionally, to assess how stochastic random initializations of k-means affect the clustering structure, we repeated it 49 times. Due to the high computational cost of this step, multiple runs of k-means, k-means++ and hk-means were limited to the comparisons between configurations with *C. chrysops* and *C. morio*.

DIAGNOSTICS

Given the number of samples of each species, the complex variation in their acoustic structure, and the many alternative clustering results, a direct inspection is logistically and intuitively infeasible. Therefore, to perform this task, we used external and internal metrics as quantitative criteria under which evaluations and comparisons could be done. The term 'external' stands for the dependence on external data, or a ground-truth, to which a clustering result can be compared and evaluated with agreement or disagreement as criteria. On the other hand, 'internal' stands for features related to the shapes, compactness, separation, and many other aspects describing how clusters relate to dissimilarity structure within the data used to build it (Kassambara 2017).

In the first step of our analysis of the diagnostics, we used internal metrics to evaluate how our clustering results fitted to the dissimilarity used to generate it. For this task we chose two of the most used metrics: Silhouette Index (GSI; Rousseeuw 1987) and Pearson's Γ (PG; Halkidi 2001), which is also referred in the literature as Normalized Hubert's Γ. With the GSI we evaluate high compactness (i.e. low intra-cluster dissimilarity) and high separation (i.e. high inter-cluster dissimilarity) as indicators of clearly defined clustering structure. The GSI is the global average of the ratio between distances of a given sample to all samples in the same cluster and all samples in other clusters, and is bound between -1 and 1. Alternatively, PG is the Pearson's correlation between the feature dissimilarity structure and a cophenetic matrix, of size N x N, where N is the number of instances or samples in the data, in which zeros are assigned to those belonging to the same cluster and one for those belonging to different clusters (Halkidi 2001). In PG, high clustering signal is interpreted as high compliance between the data dissimilarity structure and cluster memberships. Both GSI and PG were interpreted in function of the numbers of clusters, in a k-sweep strategy (Kassambara 2017; Kershenbaum *et al.* 2013, 2016; Wadewitz *et al.* 2015). The highest GSI and PG within the k-sweep could likely be found at low k, and tend to decay with increasing k. For this reason, peaks or stable regions at higher k are the ones with best adjustment to the criteria of GSI and PG. It should be noted though, that this first step of the diagnostics is limited to how different configurations perform at identifying the expected contrast between results obtained independently for C. *chrysops* and *C. morio*,

We also performed direct comparisons between the sets of labels that indicate to which cluster each signal belongs according to configuration. For this purpose, we used an external metric, Meilă's Variation of Information (hereafter MVI; Meilă 2007; package 'mcclust' v.1.0; Fritsch 2012). The MVI measures the amount of shared information between two sets of cluster labels by comparing simultaneous gain and loss of information for each case. One of the advantages of the MVI is the possibility to compare the agreement between cluster results produced with different numbers of clusters. Instead of comparing the labels to one or more ground-truths, we compared the relative MVI between all configurations, in a pairwise approach referred as clustering space. In such clustering space each instance is one among tested configurations, and distances are interpreted as the variation of information between configuration pairs. We performed a permutational analysis of variance (PERMANOVA; Anderson 2001), as implemented in the package 'vegan' (v.2.4-6; Oksanen *et al.* 2016) test if analytical decisions have location effects in the clustering space of each species.

Lastly, we propose a strategy for feature generation based in a second set of internal metrics, represented across k-sweeps (*see* Clustering Algorithms and Parameters). These metrics are: average within-cluster dissimilarity (WCD), representing how variable vocalization types are; average inter-cluster dissimilarity (BCD), representing how different vocalization types are; and WB-ratio, as the ratio between the last WCD and BCD, which indicates the relative scale of vocalization types to the full repertoire. Although not optimized to measure clustering signal as GSI and PG, this second set of metrics provide simple and accurate representations of relevant repertoire features, which can only be accessed in dissimilarities when clustering information is available. All metrics mentioned above were calculated using the R package 'fpc' 2.1-11.1 (Hennig 2015). The final set of configurations identified as the most robust ones were directly inspected in two aspects: t-Distributed Stochastic Neighbour Embedding (t-SNE; package 'tsne' v 0.1-3; Donaldson 2016; see also Maaten & Hinton 2008), to inspect the dissimilarity structure of features; and overlapped spectrograms from all vocalizations inside each cluster, with energy averaged by distance to cluster centroid. Except from sample selection and processing, our analyses were fully processed in R with computational resources from OCCAM computer (Aldinucci *et al.* 2017).

Results

METHOD EVALUATION

Although overall clustering signal was not high (GSI < 0.4; PG < 0.6), most of the 28118 clustering configurations we tested provided consistent evidence of higher clustering signal in data from *C. chrysops* than from *C. morio* (Figure 1). The type of acoustic analysis used to build the tested features (PAF, univariate or multivariate time-series) was the most influential analytical decision. It revealed large consistency between independent methods, which provide similar insights about the expected differences in their clustering signals (see Figure S1). In this perspective, GSI and PG of clustering algorithms, slightly higher in h-ward and hk-means than in PAM, were interpreted as variation in performance. Although its clustering signal was not superior, hk-means overcame the stochasticity of k-means and k-means++ and avoided potential decreases in GSI and PG with a single deterministic result. Therefore, only deterministic clustering algorithms were included in further results and diagnostics.

The clustering spaces built using MVI improved significantly our assessments about the effects of the tested configurations. Although total variation was smaller in *C. chrysops* (SS = 2031.43) than in *C. morio* (SS = 4158.6; Table 1), both clustering spaces were highly correlated (MVI; Mantel; 1000 permutations; R = 0.767; $p < 0.001$), indicating similar effects from analytical decisions in both repertoires. In apparent agreement to the patterns in GSI and PG, the most influential factors over this variation were type of acoustic analysis, as PAF, univariate time-series, and multivariate time-series (see Table 1 and Figure 2). All other potential sources of variation were significant but much smaller in the magnitude of their effects (SS in Table 1). Therefore, when represented by different types of acoustic analyses, the same vocalizations can be clustered in different ways and provide different insights about each repertoire. The only exception, in which large variation was found within a type of acoustic analysis, were the comparisons between *fs*DFC and *fs*FFC in *C. morio* (see Figure 2). Also, under constraints from the other factors referred above, increasing the number of clusters introduced variation and produced the divergent patterns in the clustering space ordination (Figure 2; also notice configurations with higher k as the most peripheral in Figure S2).

The third species *C. coeruleus*, was intermediate to the *C. chrysops* and *C. morio* in many aspects, including clustering signal and overall MVI between configurations (Figure S3, Figure S4 and Table S3). It provided an additional perspective from which analytical decisions, their caveats, and assumptions were evaluated. According to the literature, the known repertoire of *C. coeruleus* is clearly more diverse than the repertoire of *C. morio*, but *fs*FFC produced nearly identical BCD and WCD for both species (Figure 3). Given this redundancy, and the divergence in MVI between *fs*FFC and *fs*DFC in data from *C. morio*, we did not include *fs*FFC in further comparisons. The only feature space that produced higher BCD and WCD for *C. morio* than for *C. coeruleus* was *fs*LFB (Figure 3), and for this reason we chose to represent multivariate time-series exclusively by *fs*MFC. Pre-processing techniques improved clustering signal but removed and altered dimensions differently for each species. As method artifacts, BCD scale could not be considered independent of WCD scale (Figure 3), introducing bias in WB-ratio and making comparisons problematic. Although random forest coerces maximum dissimilarity to unit scale, its non-linear effects produced clusters that only partially satisfied our criteria and were complex to interpret. Even though warping amount was different between feature spaces, DTW preserved dissimilarity scales within the same features and provided BCD and WCD useful for species comparisons (Figure 3; see also PAF in Figure S5). Therefore, we identified *fs*DFC and *fs*MFC as the most suitable feature spaces for further inspection and use for species comparisons.

THE ACOUSTIC STRUCTURE OF NWJ REPERTOIRES

We found quantitative support for our hypotheses about interspecific differences concerning repertoire features. In comparisons under the same algorithm, feature, and *k*, we found remarkably stable results for all *k* equal to or larger than those at GSI and PG optima, despite the potential decrease in clustering signal. Using BCD as a measure of how different vocalizations can be in a repertoire, the highest values were from *C. chrysops*, intermediate from *C. coeruleus* and the lowest from *C. morio*. Using WCD to measure variation inside vocalization types are, the lowest values were from *C. chrysops*, intermediate from *C. morio*, and the highest from *C. coeruleus*. The last clustering feature, WB-ratio, which represents the relationship between the other two repertoire features, was lower in *C. chrysops*, intermediate in *C. coeruleus*, and higher in *C. morio*. Although the contrast between *C. chrysops* and *C. morio* was similar in all cases, *C. coeruleus* had higher BCD and WCD on *fs*MFC configurations than on *fs*DFC. Still, WB-ratio represents how diffuse or continuous acoustic variation is between vocalization types in *C. coeruleus* and *C. morio*. Visualizations of *fs*DFC and *fs*MFC feature spaces under t-SNE neighbourhood probabilities are consistent with patterns described by clustering features (Figure 4).

To illustrate the intuition behind our results and interpretations, we inspected clusters generated by PAM and h-ward from *fs*MFC. Vocalization types of *C. morio* seem, indeed, to be variants of the same harmonically rich structure, produced in variable duration and modulation patterns (Figure 5). More specifically, such modulation patterns consist of soft upward and downward modulations forming a positively skewed arch, which could be interrupted by an abrupt break in the harmonic structure, at the inflexion point. In *C. coeruleus* appears to contain a gradient resembling a slightly simplified version of the one detected in *C. morio*, lacking frequency breaks (Figure 5). Differently, though, some vocalization types from *C. coeruleus* are composed of highly entropic calls, which lack a clear harmonic structure, and have no equivalents in data from *C. morio*. At this step, it is possible to visualize how *C. chrysops* repertoire appears to be more complex than the other two (Figure 6). Although we also identified a gradient from short to long modulated calls, in this case, several other cases reveal variation in more discrete aspects. Examples of such are short guttural call types, usually repeated at least two times per utterance, and several long and harmonically rich calls, which can have a guttural portion embedded in their harmonics at varying positions.

Discussion

Clustering signal is an essential step for cluster validation, but our results using MVI and clustering features revealed other important patterns that would not be detected otherwise. Regarding clustering configurations, little variation was found between results from different clustering algorithms, with very few conditions in which one alternative outperformed the other (Guyon *et al.* 2009; Hennig & Liao 2013). Similarly to results obtained by Wadewitz *et al.* (2015) and Kershenbaum *et al.* (2013), ours point to feature selection as the most important aspect when clustering vocalizations. The alternative diagnostics represented by distances between clusters and within clusters, allowed us to demonstrate how clustering results produced using PAF are affected by pre-processing techniques. In this case, changes in scale made comparisons between repertoires seems to be problematic, even with results from the same configurations. Patterns in MVI allowed us to identify how insights from alternative descriptors can be different. The analytical decisions, which explained most MVI, were the types of acoustic analyses used to describe vocalizations, an indication of their essentially different information contents.

We demonstrated how cluster analysis, an unsupervised approach, can be effectively used to measure repertoire features. To represent repertoire features, such as acoustic variation between and within vocalization types, and their ratio in k-sweeps was indeed a more effective way to describe variation in the studied repertoires than counts of vocalization types. Besides this, we confirmed the possibility of comparing clustering results for different repertoires, observed the use of similar configurations and the methods used for acoustic analysis preserved the dissimilarity scales across independent runs. Time-series aligned by dynamic time warping satisfy this condition, so we strongly recommend its use in detriment of PAF in detriment of vocalization type counts, especially when interspecific comparisons are intended.

For this reason, we considered dominant frequency contours and Mel-Frequency Cepstral Coefficients the best alternatives to compare the repertoires used as examples here. Remarkably, despite significant differences in information content, as evident through MVI, repertoire features measured from both methods agreed with each other and with interspecific differences predicted from the available literature. Instead of suggesting the number of vocalizations in each repertoire, with support of a single value for k, our results were consistent throughout large intervals of the k-sweeps. Therefore, in this approach, assumptions involved in the choice of a single value of k are relaxed and reproducibility is improved by allowing comparisons under the same exact configurations. In practical terms, differences between repertoires could still be effectively detected when the chosen k produces low GSI or PG. However, if a single value of k must be selected, it must be higher than those in which peaks in GSI and PG were found, considering all repertoires. In our study we found this threshold to be at k = 8, according to what we found in *C. chrysops* repertoire. While GSI and PG, ultimately decay to similar levels for all repertoires when increasing the number of clusters beyond this threshold, the metrics we chose as repertoire features stabilized maintaining stable differences between species. Further exploration and a better definition of this pattern is necessary to verify if it is a limitation or an advantage.

Similarly to Gamba *et al.* (2015), we found clustering results to be less detailed than those from the available descriptions of repertoires from *C. coeruleus* and *C. chrysops* (see Hardy 1969; Anjos & Vielliard 1993; Uejima 1998). However, for *C. morio*, we found a slightly more complex and subtle pattern than the one described by Hardy (1969). An example of possible overestimation of vocalization diversity could the fourth (in h-ward) and fifth (in PAM) to the ninth cluster spectrograms of *C. chrysops* in Figure 6. Their combined spectrograms represent the most frequently recorded vocalization in the field, described by Uejima (1998) as ‘R12’ and latter referred by Brunetta & Anjos (2010) as ‘social call’. Because those samples were recorded in a single locality, this variation could likely be the product of individual or group vocal signatures. An example of possible underestimation could be the first three cluster spectrograms from h-ward, and the first three from PAM. Those examples include at least 12 short vocalization types described by Uejima (1998), which could only be properly represent because when features from series, such as the number of repetitions and interval duration, are included.

Even under controlled and robust experimental designs accounting for all sources of variation in corvid vocalizations is not a simple task (Mates *et al.* 2014; Luef *et al.* 2017). To do it, biological validation with field data, such as playback experiments, is required (see Fisher *et al.* 2013). This trend towards simplification in unsupervised results could be a consequence of unbalanced representation of potential vocalization types as consequence of how frequently different vocalization types were recorded in the field. Dealing with this issue with subsampling methods, in an unsupervised manner, would be worth exploring. Alternatively, according to Hauser (1996), human observers are prone to recognize extremes of gradients as discrete categories, even in cases in which quantitative analysis or the vocalizing species do not (see also Gamba *et al.* 2015). It appears to be a perceptual bias towards rare features, making observers prone to define vocalization types from atypical or rare vocalizations. This question is also worth exploring because classification models used to classify new data often rely in human observers to label training data.

There is a myriad of methods of clustering and classification methods, which application goes beyond Bioacoustics (Hennig et al. 2016). However, application specific recommendations must be updated. As animal sound collections grow, a reproducible method capable of dealing with large and complex datasets becomes increasingly important. New techniques allow insight about patterns never accessed before, but many not-so-new methods have been underexplored for this purpose. As we demonstrate in this study, tools already available at software packages for cluster analysis could be combined in new ways to solve old problems. Therefore, we encourage thorough evaluation of the methods included in this study, such as the Dynamic Time Warping algorithm, and acoustic analyses beyond this scope. The process of sound recording, storage, and analysis inevitably lose or modify information from the original sound, and each technique has its drawbacks (Araya-Salas *et al.* 2017). So, minimizing the amount of decisions and automating processes is essential for improving reproducibility and assessing new patterns.

Disclosure Statement

The authors declare no conflicts of interest.

Acknowledgements

For valuable insights on this study, we thank C. De Gregorio, D. Valente, E. Cisselo, G. B. di Panfilo, J. R. P. Adelino, M. Araya-Salas, M. R. Lima, and V. Torti. Field sampling was conducted with authorization of private land owners. Sampling in Mata dos Godoy State Park was authorized by IAP – Instituto Ambiental do Paraná.

Funding

This study was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and by CONFAP and Fundação Araucária, through the program Mobility CONFAP Italy.

References

Aldinucci, M., Bagnasco, S., Lusso, S., Pasteris, P., Rabellino, S., & Vallero, S. (2017). OCCAM: A flexible, multi-purpose and extendable HPC cluster. *Journal of Physics: Conference Series*, 898. DOI 10.1088/1742-6596/898/8/082039

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26, 32-46. DOI 10.1046/j.1442-9993.2001.01070.x

Anjos, L. Dos, & Vielliard, J. M. E. (1993). Repertoire of the acoustic communication of the azure jay *Cyanocorax caeruleus* (Vieillot) (Aves, Corvidae). *Revista Brasileira de Zoologia*, 10, 657-664. DOI 10.1590/S0101-81751993000400011

Anjos, L. Dos. (1991). O ciclo anual de *Cyanocorax caeruleus* em floresta de araucária (Passeriformes: Corvidae). *Ararajuba*, 2, 19-23.

Araya-Salas, M., & Smith-Vidaurre, G. (2017). warbleR: an r package to streamline analysis of animal acoustic signals. *Methods in Ecology and Evolution*, 8, 184-191. DOI 10.1111/2041-210X.12624

Araya-Salas, M., Smith-Vidaurre, G., & Webster, M. (2017). Assessing the effect of sound file compression and background noise on measures of acoustic signal structure. *Bioacoustics*, 28, 57-73. DOI 10.1080/09524622.2017.1396498

Arthur, D., & Vassilvitskii, S. (2007). K-Means++: the Advantages of Careful Seeding. *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 8, 1027-1025. DOI 10.1145/1283383.1283494

Boesing, A. L., & Anjos, L. Dos. (2012). The Azure Jay may reproduce in plantations of *Araucaria angustifolia* in southern Brazil. *Bird Conservation International*, 22, 205-212. DOI 10.1017/S0959270911000529

Bonaccorso, E., Peterson, A. T., Navarro-Sigüenza, A. G., & Fleischer, R. C. (2010). Molecular systematics and evolution of the *Cyanocorax* jays. *Molecular Phylogenetics and Evolution*, 54, 897-909. DOI 10.1016/j.ympev.2009.11.014

Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5-32. DOI 10.1023/a:1010933404324

Brunetta, B., & Anjos, L. Dos (2010). Variações na distribuição espacial e no grito social da gralha-picaça (*Cyanocorax chrysops*) na Mata Atlântica, Sul do Brasil. *Ornitologia Neotropical*, 21, 203-213.

Burghardt, G. M., Bartmess-Levasseur, J. N., Browning, S. A., Morrison, K. E., Stec, C. L., Zachau, C. E., & Freeberg, T. M. (2012). Perspectives - Minimizing Observer Bias in Behavioral Studies: A Review and Recommendations. *Ethology*, 118, 511-517. DOI 10.1111/j.1439-0310.2012.02040.x

Coppinger, B., Cannistraci, R. A., Karaman, F., Kyle, S. C., Hobson, E. A., Freeberg, T. M., & Hay, J. F. (2017). Studying audience effects in animals: what we can learn from human language research. *Animal Behaviour*, 124, 161-165. DOI 10.1016/j.anbehav.2016.12.020

Daniel Meliza, C., Keen, S. C., & Rubenstein, D. R. (2013). Pitch- and spectral-based dynamic time warping methods for comparing field recordings of harmonic avian vocalizations. *The Journal of the Acoustical Society of America*, 134, 1407-1415. DOI 10.1121/1.4812269

Davis, S. B., & Mermelstein, P. (1980). Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences. *IEEE Transactions on Acoustics, Speech, and Signal Processing*. DOI 10.1109/TASSP.1980.1163420

Deecke, V. B., & Janik, V. M. (2005). Automated categorization of bioacoustic signals: Avoiding perceptual pitfalls. *The Journal of the Acoustical Society of America*, 117, 2470-2470. DOI 10.1121/1.4787454

Donaldson, J. (2016). tsne: T-Distributed Stochastic Neighbor Embedding for R (t-SNE). R package version 0.1-3. URL https://CRAN.R-project.org/package=tsne

Elie, J. E., Soula, H. A., Mathevon, N., & Vignal, C. (2011). Dynamics of communal vocalizations in a social songbird, the zebra finch (*Taeniopygia guttata*). *The Journal of the Acoustical Society of America*, 129, 4037-4046. DOI 10.1121/1.3570959

Fischer, J., Noser, R., & Hammerschmidt, K. (2013). Bioacoustic field research: A primer to acoustic analyses and playback experiments with primates. *American Journal of Primatology*, 75, 643-663. DOI 10.1002/ajp.22153

Florentin, J., Dutoit, T., & Verlinden, O. (2016). Identification of European woodpecker species in audio recordings from their drumming rolls. *Ecological Informatics*, 35, 61-70. DOI 10.1016/j.ecoinf.2016.08.006

Fritsch, A., 2009. mcclust: Process an MCMC Sample of Clusterings. http://cran.r-project.org/package=mcclust

Gamba, M., Friard, O., Riondato, I., Righini, R., Colombo, C., Miaretsoa, L., Torti, V., Nadhurou, B. & Giacoma, C. (2015). Comparative Analysis of the Vocal Repertoire of *Eulemur*: A Dynamic Time Warping Approach. *International Journal of Primatology*, 36, 894-910. DOI 10.1007/s10764-015-9861-1

Giorgino, T. (2009). Computing and Visualizing Dynamic Time Warping Alignments in R: The dtw Package. *Journal of Statistical Software*, 31, 1-24. DOI 10.18637/jss.v031.i07

Große Ruse, M., Hasselquist, D., Hansson, B., Tarka, M., & Sandsten, M. (2016). Automated analysis of song structure in complex birdsongs. *Animal Behaviour*, 112, 39-51. DOI 10.1016/j.anbehav.2015.11.013

Guyon, I., von Luxburg, U. & Williamson, R.C.. (2009). Clustering: science or art? *NIPS 2009 Workshop on Clustering Theory*, Vancouver, Canada.

Guyon, I., Von Luxburg, U., & Williamson, R. (2009). *Clustering: science or art*. NIPS 2009 workshop on clustering theory.

Hale, A. M., Williams, D. A., & Rabenold, K. N. (2003). Territoriality and Neighbor Assessment in Brown Jays (*Cyanocorax morio*) in Costa Rica. *The Auk*, 120, 446-456. DOI 10.2307/4090197

Halkidi, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 173, 107-145. DOI 10.1023/A:1012801612483

Hardy, J. (1969). A taxonomic revision of the New World jays. *The Condor*, 71, 360-375. DOI 10.2307/1365735

Hauser, M. D. (1996). *The evolution of communication*. Cambridge, MA: MIT Press.

Hennig, C. (2018). fpc: Flexible Procedures for Clustering. R package version 2.1-11.1. URL https://CRAN.R-project.org/package=fpc

Hennig, C., & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society. Series C: Applied Statistics*, 62, 309-369. DOI 10.1111/j.1467-9876.2012.01066.x

Hennig, C., Meilă, M., Murtagh, F., & Rocci, R. (2016). *Handbook of cluster analysis*. DOI 10.1111/rssa.12219

Husson, F., Le, S., & Pages, J. (2010). *Exploratory multivariate analysis by example using R*.

Jensen, K. K., & Klokker, S. (2006). Hearing sensitivity and critical ratios of hooded crows (*Corvus corone cornix*). *The Journal of the Acoustical Society of America*, 119, 1269. DOI 10.1121/1.2159431

Josse, J., & Husson, F. (2012). Selecting the number of components in principal component analysis using cross-validation approximations. *Computational Statistics and Data Analysis*, 56, 1869-1879. DOI 10.1016/j.csda.2011.11.012

Kahl, S., Wilhelm-Stein, T., Hussein, H., Klinck, H., Kowerko, D., Ritter, M., & Eibl, M. (2017). Large-scale bird sound classification using convolutional neural networks. *CEUR Workshop Proceedings*, *1866*.

Kassambara, A. & Mundt, F., (2017). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.5. URL https://CRAN.R-project.org/package=factoextra

Kassambara, A. (2017). *Practical guide to cluster analysis in R*. STHDA.

Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M., & Farnsworth, A. (2014). A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). *Ecological Informatics*, 21, 25-33. DOI 10.1016/j.ecoinf.2014.01.001

Kershenbaum, A., Sayigh, L. S., & Janik, V. M. (2013). The Encoding of Individual Identity in Dolphin Signature Whistles: How Much Information Is Needed? *PLoS ONE*, 8, 1-7. DOI 10.1371/journal.pone.0077671

Krull, C. R., Ranjard, L., Landers, T. J., Ismar, S. M. H., Matthews, J. L., & Hauber, M. E. (2012). Analyses of sex and individual differences in vocalizations of Australasian gannets using a dynamic time warping algorithm. *The Journal of the Acoustical Society of America*, 132, 1189-1198. DOI 10.1121/1.4734237

Lawton, M. F., & Guindon, C. F. (1981). Flock composition , breeding success , and learning in the Brown jay. *The Condor*, 83, 27-33.

Ligges, U., Krey, S., Mersmann, O., & Schnackenberg, S. (2018). tuneR: Analysis of Music and Speech. URL: https://CRAN.R-project.org/package=tuneR

Luef, E. M., Ter Maat, A., & Pika, S. (2017). Vocal similarity in long-distance and short-distance vocalizations in raven pairs (*Corvus corax*) in captivity. *Behavioural Processes*, 142, 1-7. DOI 10.1016/j.beproc.2017.05.013

Maaten, L. van der, & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9, 2579-2605.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K.(2018). cluster: Cluster Analysis Basics and Extensions. R package version 2.0.7-1.

Mates, E. A., Tarter, R. R., Ha, J. C., Clark, A. B., & McGowan, K. J. (2014). Acoustic profiling in a complexly social species, the American crow: caws encode information on caller sex, identity and behavioural context. *Bioacoustics*, 24, 63-80. DOI 10.1080/09524622.2014.933446

Meilă, M. (2007). Comparing clusterings-an information based distance. *Journal of Multivariate Analysis*, *98*(5), 873-895. DOI 10.1016/j.jmva.2006.11.013

Mouselimis, L. (2017). ClusterR: Gaussian Mixture Models, K-Means, Mini-Batch-Kmeans and K-Medoids Clustering. R package version 1.0.6. URL https://CRAN.Rproject.org/package=ClusterR

Murtagh, F., & Legendre, P. (2014). Ward's Hierarchical Agglomerative Clustering Method : Which Algorithms Implement Ward's Criterion? *Journal of Classification*, 31, 274-295. DOI 10.1007/s00357-

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., & O'Hara, R. (2016). Vegan: community ecology package. *R Package*. https://cran.r-project.org/package=vegan

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ratanamahatana, C. A., & Keogh, E. (2004). Everything you know about Dynamic Time Warping is Wrong. *Third workshop on mining temporal and sequential data*, 32. Citeseer.

Reinert, B. L., & Bornschein, M. R. (1998). Alimentação da gralha-azul (*Cyanocorax caeruleus*, Corvidae). *Ornitología Neotropical*, 9, 213-217.

Riondato, I., Cissello, E., Papale, E., Friard, O., Gamba, M., & Giacoma, C. (2017). Unsupervised Acoustic Analysis of the Vocal Repertoire of the Gray-Shanked Douc Langur (*Pygathrix cinerea*). *Journal of Computational Acoustics*, 25, 1750018. DOI 10.1142/S0218396X17500187

Rosa, G. L. M., Ellis, J. M., Bonaccorso, E., & Dos Anjos, L. (2016). Friend or foe? Social system influences the allocation of signals across functional categories in the repertoires of the New World jays. *Behaviour*, 153, 467-524. DOI 10.1163/1568539X-00003360

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65. DOI 10.1016/0377-0427(87)90125-7

Sarda-Espinosa, A. (2017). *Comparing Time-Series Clustering Algorithms in R Using the dtwclust Package*. URL https://cran.r-project.org/web/packages/dtwclust/vignettes/dtwclust.pdf

Somervuo, P. (2018). Time-frequency warping of spectrograms applied to bird sound analyses. *Bioacoustics*, 28, 257-268. DOI 10.1080/09524622.2018.1431958

Uejima, A. M. K. (1998). Ecologia da gralha-picaça, *Cyanocorax chrysops* (Vieillot, 1818) (Passeriformes: Corvidae) em três áreas ao longo da Bacia do Rio Tibagi, Estado do Paraná, Brasil. Thesis, Universidade Federal do Paraná.

Uejima, A. M. K., Boesing, A. L., & Anjos, L. Dos. (2012). Breeding and Foraging Variation of the Plush-Crested Jay (*Cyanocorax chrysops*) in the Brazilian Atlantic Forest. *The Wilson Journal of Ornithology*, 124, 87-95. DOI 10.1676/11-027.1

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203-220. DOI 10.1016/j.anbehav.2016.12.005

Wadewitz, P., Hammerschmidt, K., Battaglia, D., Witt, A., Wolf, F., & Fischer, J. (2015). Characterizing vocal repertoires - Hard vs. Soft classification approaches. *PLoS ONE*, 10, e0125785. DOI 10.1371/journal.pone.0125785

Williams, D. A. (2004). Female control of reproductive skew in cooperatively breeding brown jays (*Cyanocorax morio*). *Behavioral Ecology and Sociobiology*, 55, 370-380. DOI 10.1007/s00265-003-0728-7

Williams, D. a., & Hale, A. M. (2006). Helper Effects on Offspring Production in Cooperatively Breeding Brown Jays (*Cyanocorax morio*). *The Auk*, 123, 847-857. DOI 10.1642/0004-8038(2006)123[847:HEOOPI]2.0.CO;2

Williams, D. A., Lawton, M. F., & Lawton, R. O. (1994). Population growth, range expansion, and competition in the cooperatively breeding brown jay, *Cyanocorax morio*. *Animal Behaviour*, 48, 309-322. DOI 10.1006/anbe.1994.1244

Xie, J., Towsey, M., Zhang, J., & Roe, P. (2018). Frog call classification: a survey. *Artificial Intelligence Review*, 49, 375-391. DOI 10.1007/s10462-016-9529-z

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3-14. DOI 10.1111/j.2041-210X.2009.00001.x

Tables

Table 1. Permutational Analysis of Variance (PERMANOVA; Anderson, 2001) of clustering spaces based on Meilă's Variation of Information (MVI) for clustering results from *Cyanocorax chrysops* and *Cyanocorax morio*, to identify the magnitude effect of tested analytical decisions. As factors related to the tested configurations in this test (n = 816) are included types of acoustic analysis as predefined acoustic features (PAF), which includes *fs*PAF$_{A}$, *fs*PAF$_{B}$, *fs*PAF$_{pcB}$ and *fs*PAF$_{RF}$; univariate time-series, which include *fs*DFC and *fs*FFC; multivariate time-series, which include only *fs*MFC; clustered trough algorithms with h-ward, hk-means and PAM, and the full k-sweeps, from 2 to 35 clusters.

Cyanocorax chrysops (1000 permutations)					
Analytical Decisions	**df**	**SS**	**R2**	**F**	**Pr(>F)**
Acoustic analysis	2	1400.7	0.6384	3061.2	**<0.001**
Feature space	5	69.13	0.0315	60.432	**<0.001**
Clustering algorithm	2	15.11	0.0069	33.028	**<0.001**
Number of clusters (k)	33	532.23	0.2426	70.497	**<0.001**
residual	773	176.85	0.0806		
total	815	2194	1		

Cyanocorax morio (1000 permutations)					
Analytical Decisions	**df**	**SS**	**R2**	**F**	**Pr(>F)**
Acoustic analysis	2	3384	0.7466	3910.7	**<0.001**
Feature space	5	706.4	0.1559	326.54	**<0.001**
Clustering algorithm	2	9.9	0.0022	11.409	**<0.001**
Number of clusters (k)	33	97.9	0.0216	6.8593	**<0.001**
residual	773	334.4	0.0738		
total	815	4532.7	1		

Figures

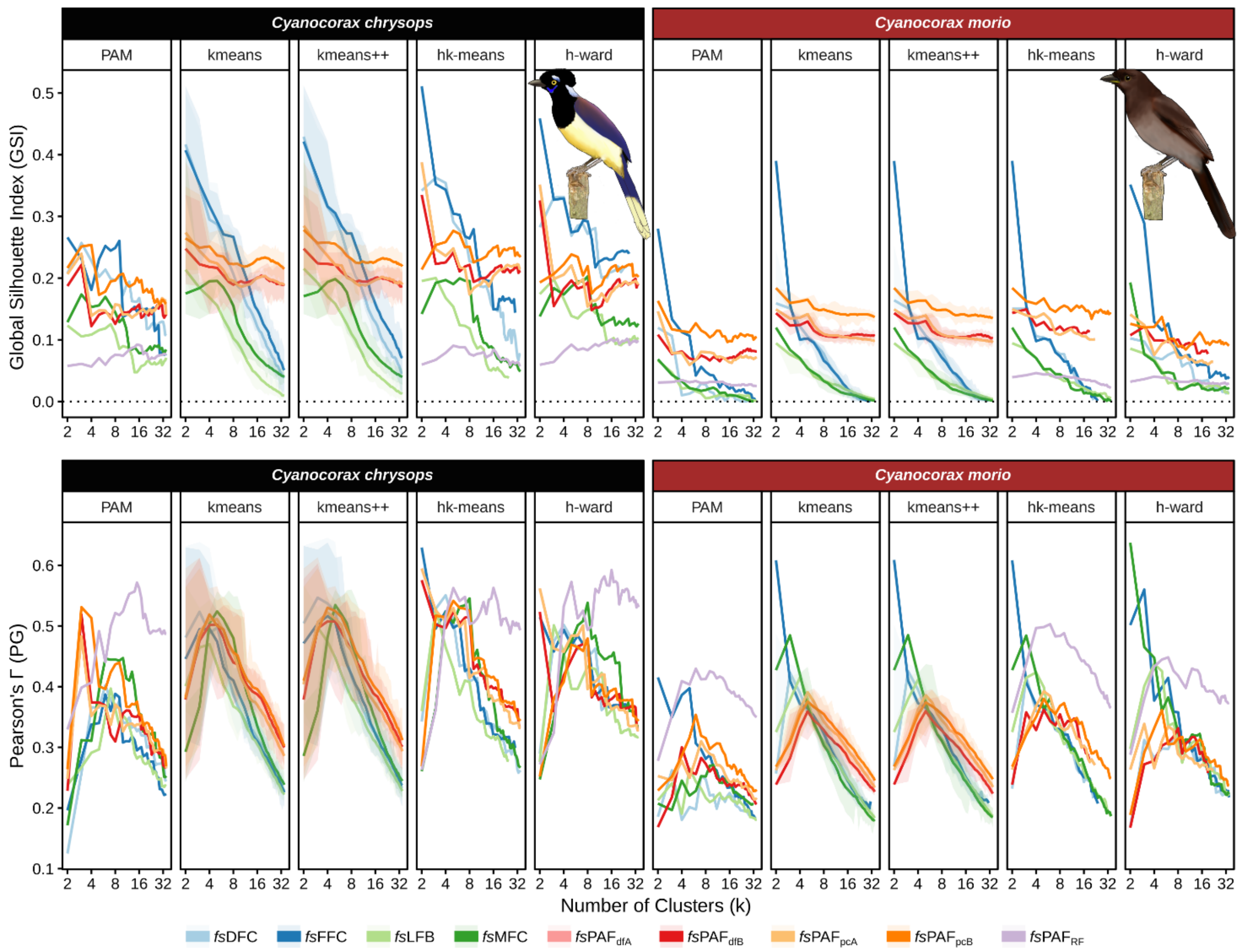


Figure 1. Global Silhouette Index (GSI) and Pearson's Γ (PG) used in the evaluation of the clustering signal evaluation using vocalization data from *Cyanocorax chrysops* and *Cyanocorax morio*. The tested configurations were based in k-sweeps from 2 to 35 clusters, represented in log scale, different acoustic features are represented by different colours, and each facet a clustering algorithm. Bands in metrics for k-means and k-means++ represent the stochastic variation we found in 50 independent runs. In both metrics, larger values correspond to higher clustering signal. Breaks in the k-sweeps represent failed configurations, such as negative silhouettes or those that generated clusters with only 5 samples or less. The feature spaces are: dominant frequency contours (*fs*DFC), fundamental frequency contours (*fs*FFC), linear frequency bins (*fs*LFB), mel-frequency cepstral coefficients (*fs*MFC), predefined acoustic features ($fsPAF_A$), predefined acoustic

features without variables with VIF superior to 10 (*fs*PAF_B), principal components of *fs*PAF_A (*fs*PAF_{pcA}), principal components of *fs*PAF_B (*fs*PAF_{pcB}), and random forest dissimilarity from the same predefined acoustic features (*fs*PAF_{RF}). Warm colours represent predefined acoustic features (PAF), tones of blue represent univariate time-series and green represent multivariate time-series.

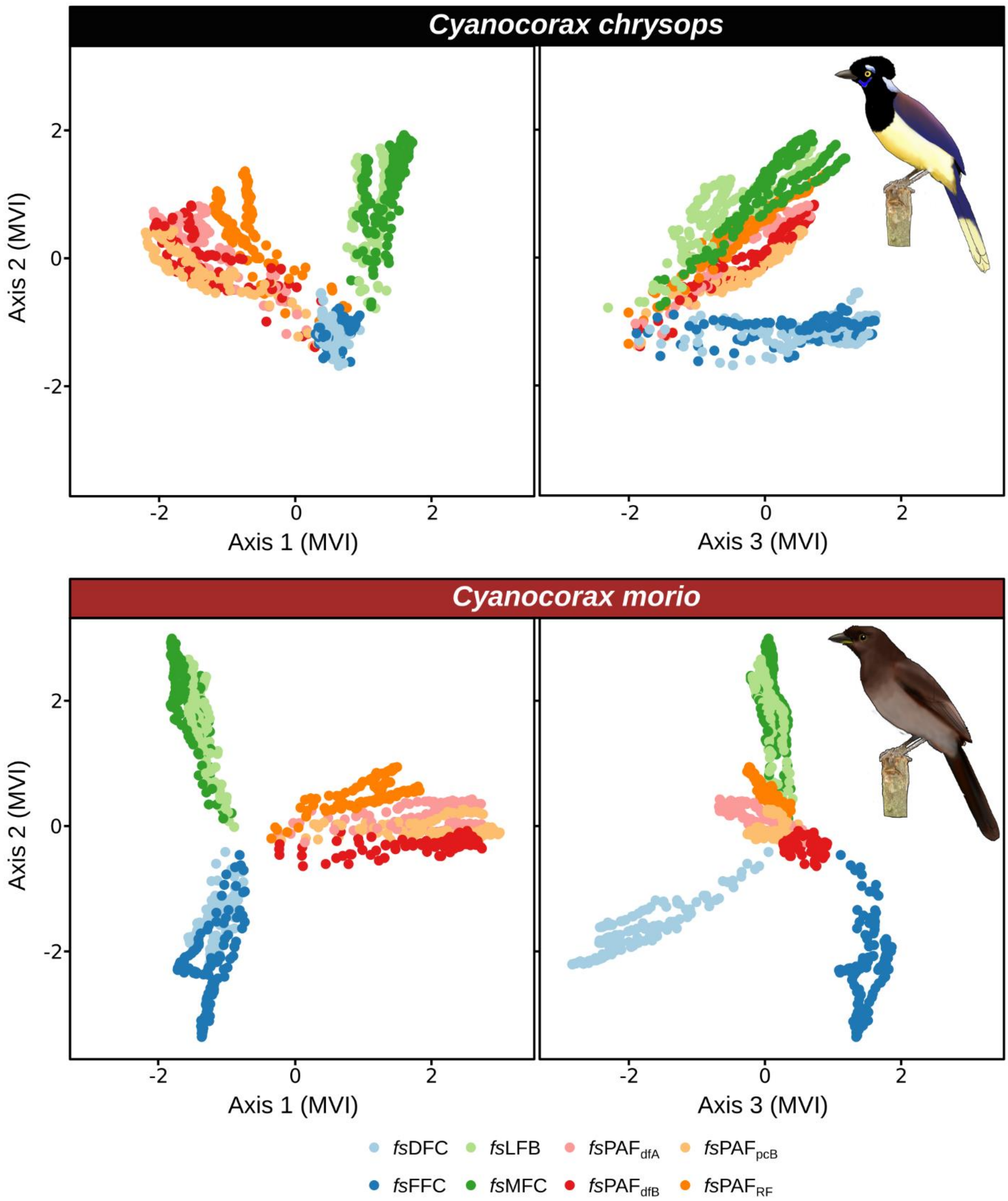


Figure 2. Ordination of pairwise Meilă's Variation of Information (MVI; Meilă, 2007) between labels obtained for *Cyanocorax chrysops* and *Cyanocorax morio* vocalizations from the tested clustering configurations. Configurations include k-sweeps of 2 ≤ k ≤ 35, and different colours represent the feature spaces: dominant frequency contours (*fs*DFC), fundamental frequency contours (*fs*FFC), linear frequency bins (*fs*LFB), mel-cepstral coefficients (*fs*MFC), predefined acoustic features (*fs*PAF$_{A}$), predefined acoustic features without

variables with VIF superior to 10 (*fs*PAF_B), principal components of *fs*PAF_A (*fs*PAF_{pcA}), principal components of *fs*PAF_B (*fs*PAF_{pcB}), and random forest dissimilarity from the same predefined acoustic features (*fs*PAF_{RF}). Warm colours represent predefined acoustic features (PAF), tones of blue represent univariate time-series and green represent multivariate time-series. Scales are the same in all panels.

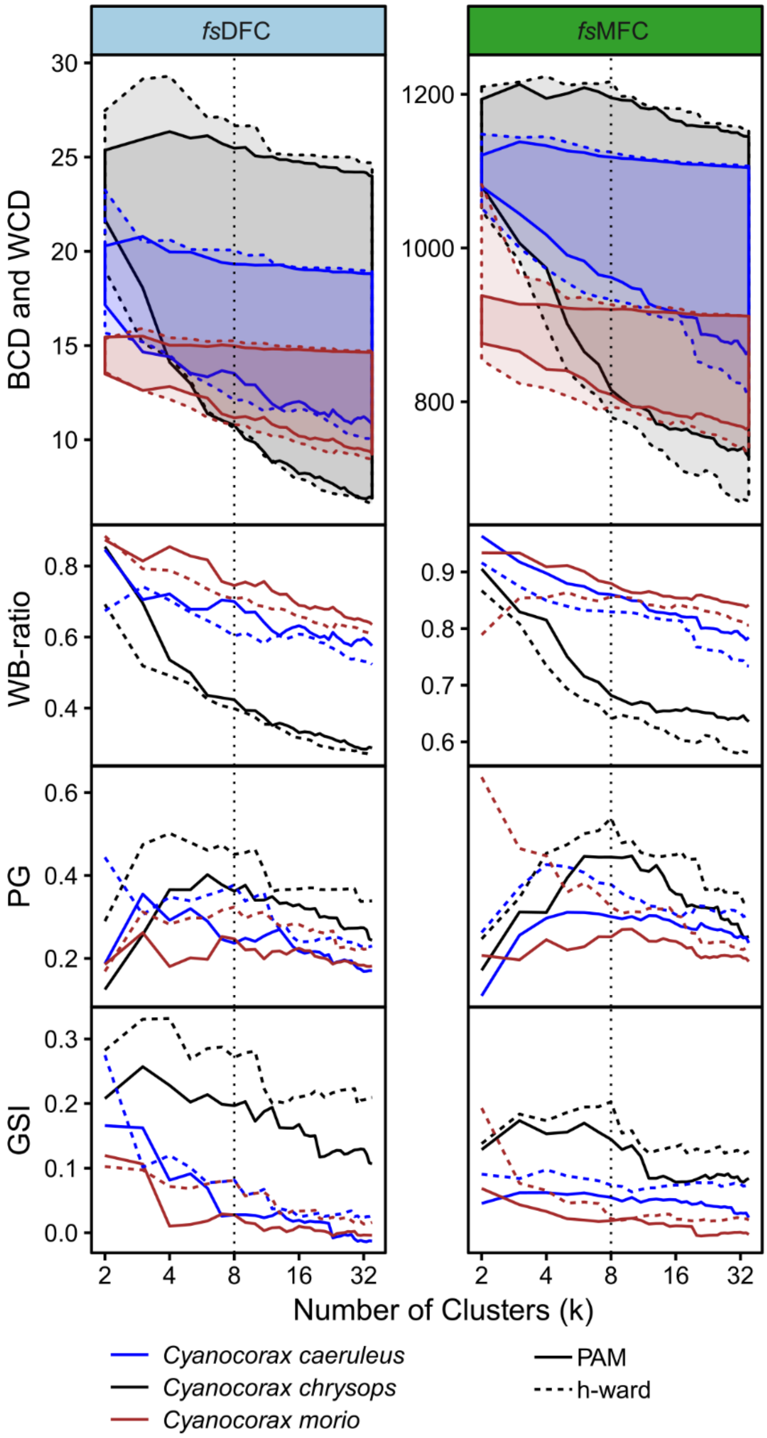


Figure 3. Clustering features of *Cyanocorax coeruleus* (blue), *Cyanocorax chrysops* (black) and *Cyanocorax morio* (brown), produced by algorithms PAM (solid lines) and h-ward (dashed lines), using *fs*DFC and *fs*MFC to represent their vocalizations in k-sweeps from 2 to 35 clusters, represented in log scale. To illustrate interspecific differences, dissimilarities between clusters (BCD) and within clusters (WCD) are represented as the upper and lower limits of shaded bands. The ratio between those clustering features is

represented below as WB-ratio, along with clustering diagnostics Global Silhouette index (GSI) and Pearson's Γ (PG) included as visual reference. With a dotted line we represented $k \geq 8$ as the interval of interest, providing stable results for interspecific comparisons.

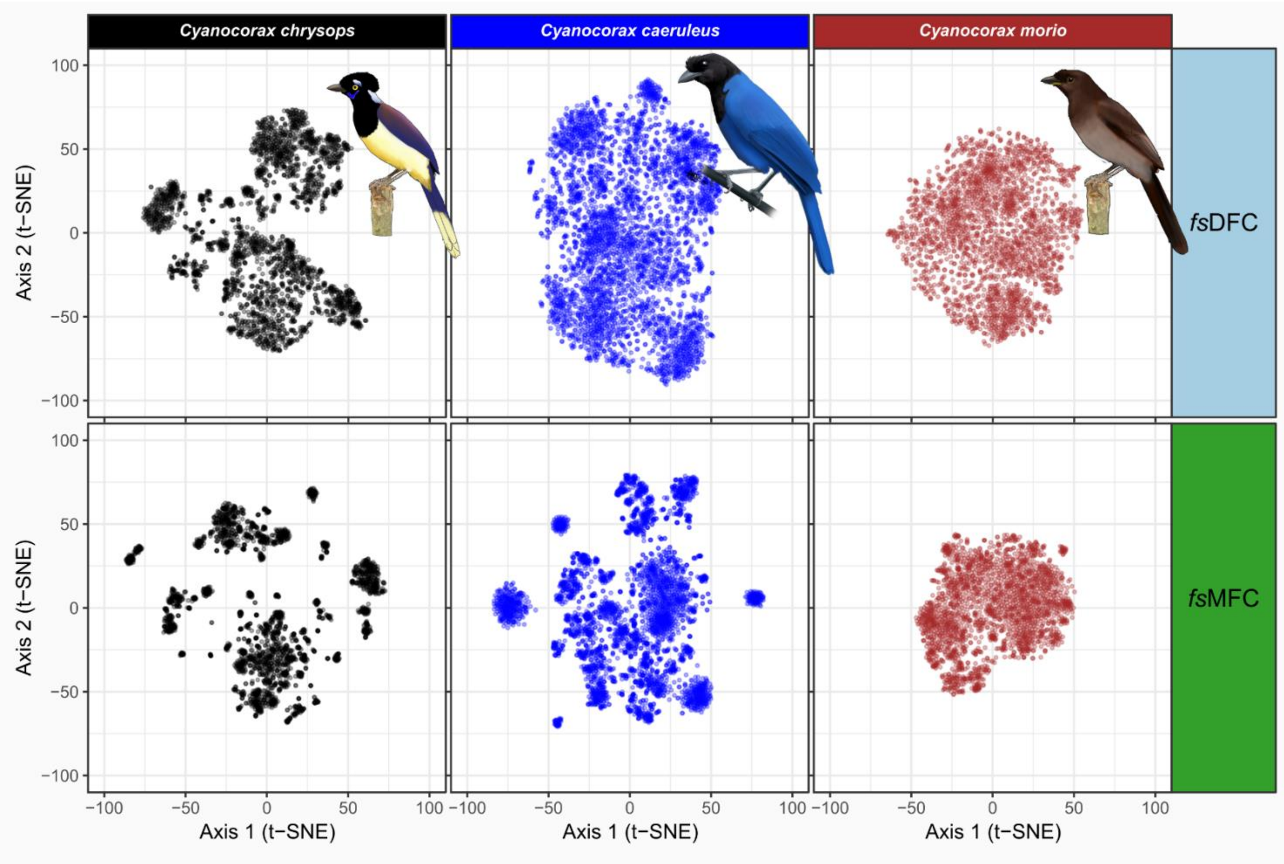


Figure 4. Two-dimensional t-SNE maps of vocalizations from *Cyanocorax coeruleus*, *Cyanocorax chrysops* and *Cyanocorax morio*, calculated from *fs*DFC and *fs*MFC (stable at 2500 iterations; perplexities = 50). Probabilistic mappings distances between and within apparent clusters are not represented linearly.

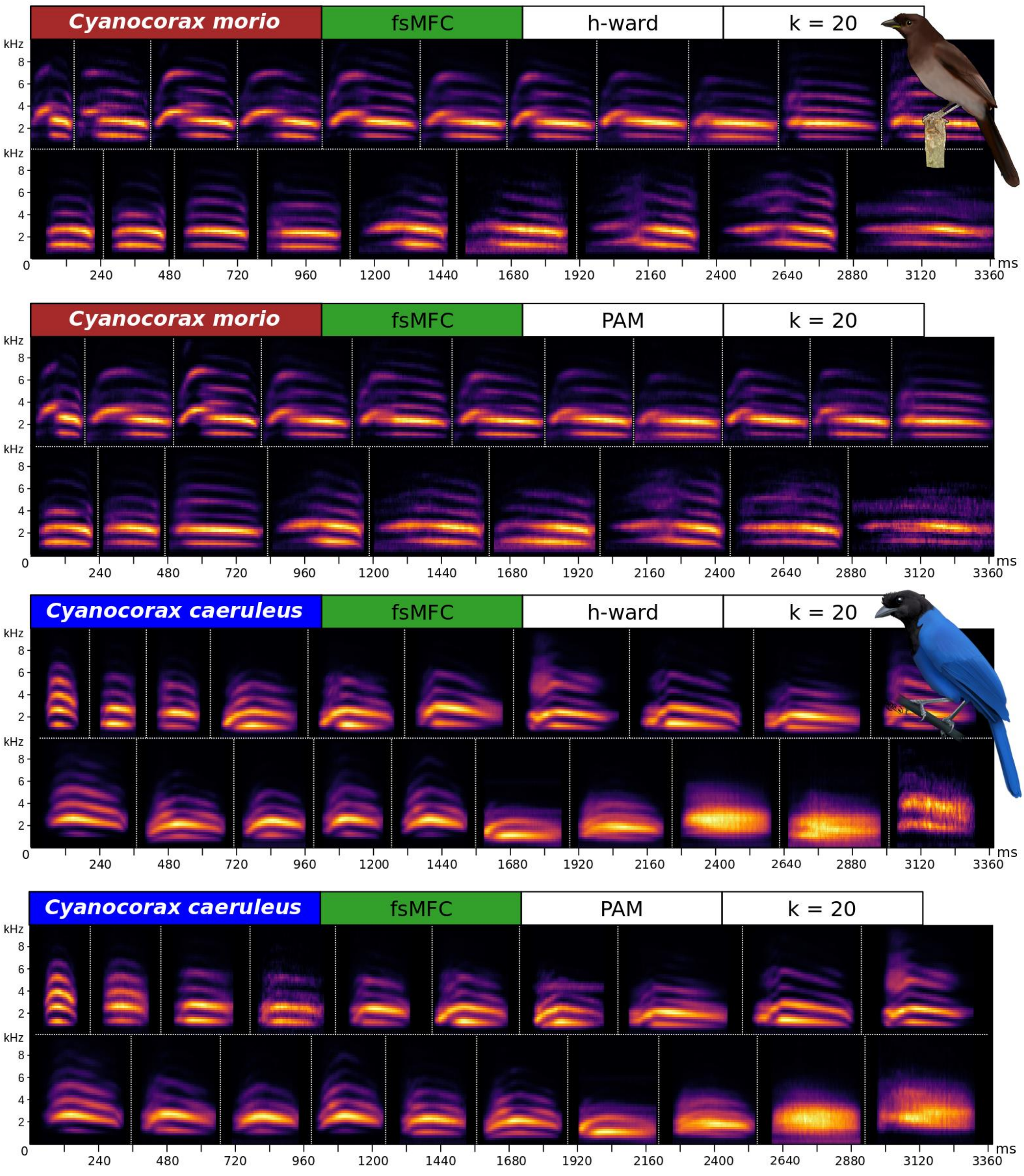


Figure 5. Combined spectrograms (f = 44100; wl = 256; ovlp = 90%) from all vocalizations within each cluster or vocalization type from *Cyanocorax coeruleus* and *Cyanocorax morio*. Contributions of each individual sample to energy and duration are weighted by their silhouettes within results from each configuration. Configurations represented 20 clusters found in fsMFC, trough h-ward and PAM algorithms.

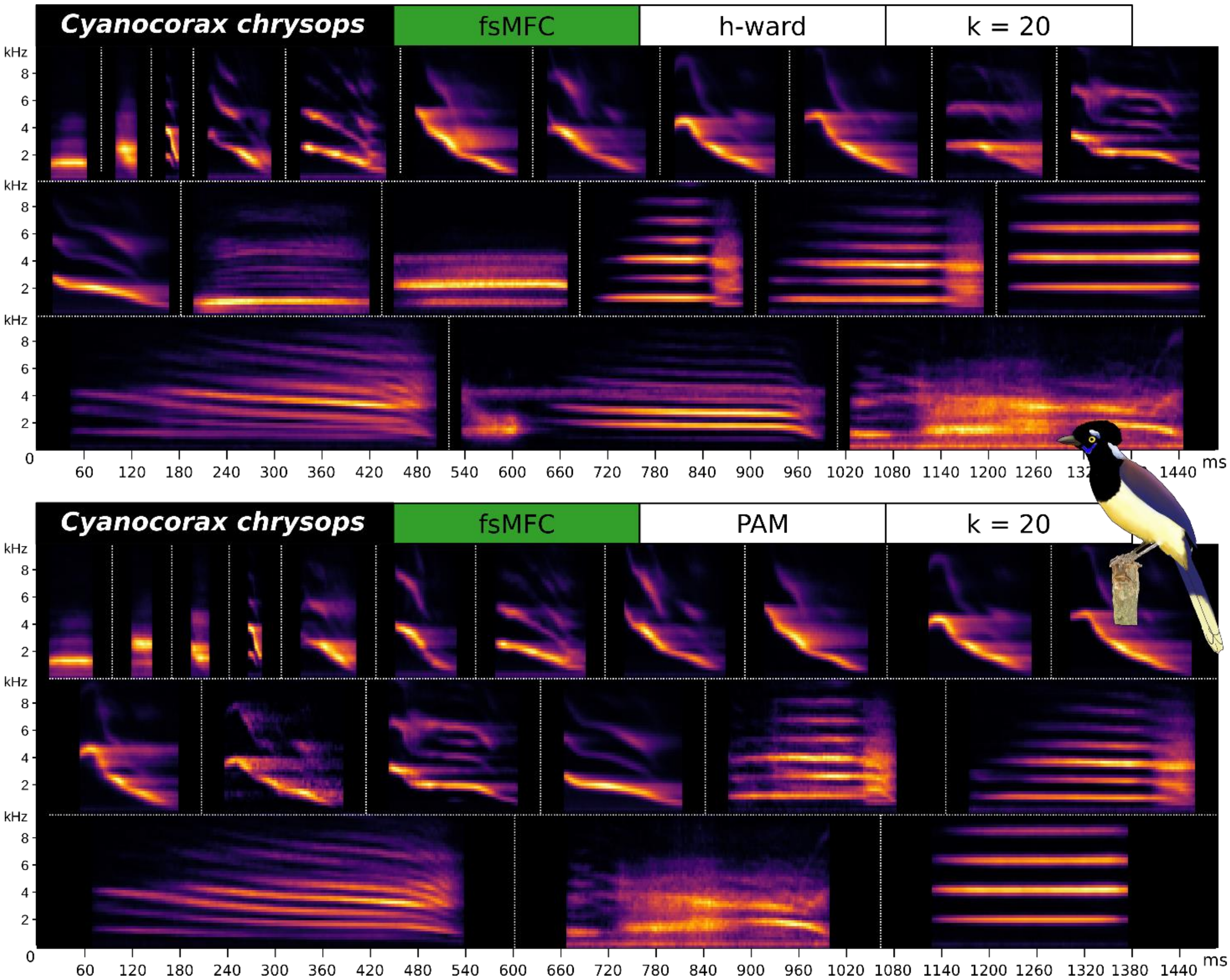


Figure 6. Combined spectrograms (f = 44100; wl = 256; ovlp = 90%) from all vocalizations within each cluster or vocalization type from *Cyanocorax chrysops*. Contributions of each individual sample to energy and duration are weighted by their silhouettes within results from each configuration. Configurations represented 20 clusters found in fsMFC, trough h-ward and PAM algorithms.

Supplementary Material

Table S1. Analytical decisions concerning the production of each feature space used as descriptor in clustering algorithms. The colour code of feature spaces is consistent across all figures.

Type of Acoustic Analysis	Acoustic Feature	Pre-processing		Feature Space
Predefined Acoustic Features (PAF)	29 acoustic variables (Table S2)	(X = 0; sd = 1)		*fs*PAF$_{A}$
			VIF	*fs*PAF$_{B}$
			PCA	*fs*PAF$_{pcA}$
			VIF + PCA	*fs*PAF$_{pcB}$
		Random Forest		*fs*PAF$_{RF}$
Univariate time-series	Fundamental Frequency Contour (FFC)	Dynamic Time Warping (DTW)		*fs*FFC
	Dominant Frequency Contour (DFC)			*fs*DFC
Multivariate time-series	Linear Frequency Bins (LFB)			*fs*LFB
	Mel-frequency Cepstral Coefficients (MFCC)			*fs*MFC

Table S2. Multivariate set of Predefined Acoustic Variables (PAF), as implemented and described in the vignette of function ‘specan’ in the package ‘warbleR’ v.1.1.12 (Araya-Salas and Smith-Vidaurre, 2017).

Variable	Unit	Definition
duration	s	Length of the signal
meanfreq	Hz	Weighted average of frequency by amplitude
sd	Hz	standard deviation of frequency weighted by amplitude
freq.median	Hz	The frequency at which the signal is divided in two frequency intervals of equal energy
freq.Q25	Hz	The frequency at which the signal is divided in two frequency intervals of 25% and 75% energy respectively
freq.Q75	Hz	The frequency at which the signal is divided in two frequency intervals of 75% and 25% energy respectively
freq.IQR	Hz	Frequency range between ‘freq.Q25’ and ‘freq.Q75’
time.median	s	The time at which the signal is divided in two time intervals of equal energy
time.Q25	s	The time at which the signal is divided in two time intervals of 25% and 75% energy respectively
time.Q75	s	The time at which the signal is divided in two time intervals of 75% and 25% energy respectively
time.IQR	s	Time range between ‘time.Q25’ and ‘time.Q75’
skew	-	Asymmetry of the spectrum
kurt	-	Peakedness of the spectrum
sp.ent	-	Energy distribution of the frequency spectrum (pure tone ~ 0; noisy ~ 1)
time.ent	-	Energy distribution on the time envelope (pure tone ~ 0; noisy ~ 1)
entropy	-	Spectral entropy
sfm	-	Spectral flatness (pure tone ~ 0; noisy ~ 1)
meanfun	Hz	Average of fundamental frequency measured across the acoustic signal
minfun	Hz	Minimum fundamental frequency measured across the acoustic signal
maxfun	Hz	Maximum fundamental frequency measured across the acoustic signal
meandom	Hz	Average of dominant frequency measured across the acoustic signal
mindom	Hz	Minimum of dominant frequency measured across the acoustic signal
maxdom	Hz	Maximum of dominant frequency measured across the acoustic signal
dfrange	Hz	Range of dominant frequency measured across the acoustic signal
modindx	-	Modulation index, calculated as the cumulative absolute difference between adjacent measurements of dominant frequencies divided by the dominant frequency range (1 means the signals is not modulated)
startdom	Hz	Dominant frequency measurement at the start of the signal
enddom	Hz	Dominant frequency measurement at the end of the signal
dfslope	kHz/s	Slope of the change in dominant frequency through time
peakf	Hz	Frequency with the highest energy

Table S3. Permutational Analysis of Variance (PERMANOVA; Anderson, 2001) of clustering spaces based on Meilă's Variation of Information (MVI) for clustering results from *Cyanocorax coeruleus*, to identify the effect magnitude of tested analytical decisions. As factors related to the tested configurations in this test (n = 816) are included types of acoustic analysis as predefined acoustic features (PAF), which includes *fs*PAF$_{A}$, *fs*PAF$_{B}$, *fs*PAF$_{pcB}$ and *fs*PAF$_{RF}$; univariate time-series, which include *fs*DFC and *fs*FFC; multivariate time-series, which include only *fs*MFC; clustered trough algorithms with h-ward, hk-means and PAM, and the full k-sweeps, from 2 to 35 clusters.

Cyanocorax coeruleus (1000 permutations)					
Analytical Decisions	**DF**	**SS**	**R^2**	**F**	***P*r(>F)**
Acoustic analysis	2	1962.99	0.70293	2923.314	**<0.001**
Feature space	4	94.85	0.03397	70.628	**<0.001**
Clustering algorithm	2	34.89	0.01249	51.957	**<0.001**
Number of clusters (k)	33	474.24	0.16982	42.803	**<0.001**
residual	672	225.62	0.08079		
total	713	2792.59	1		

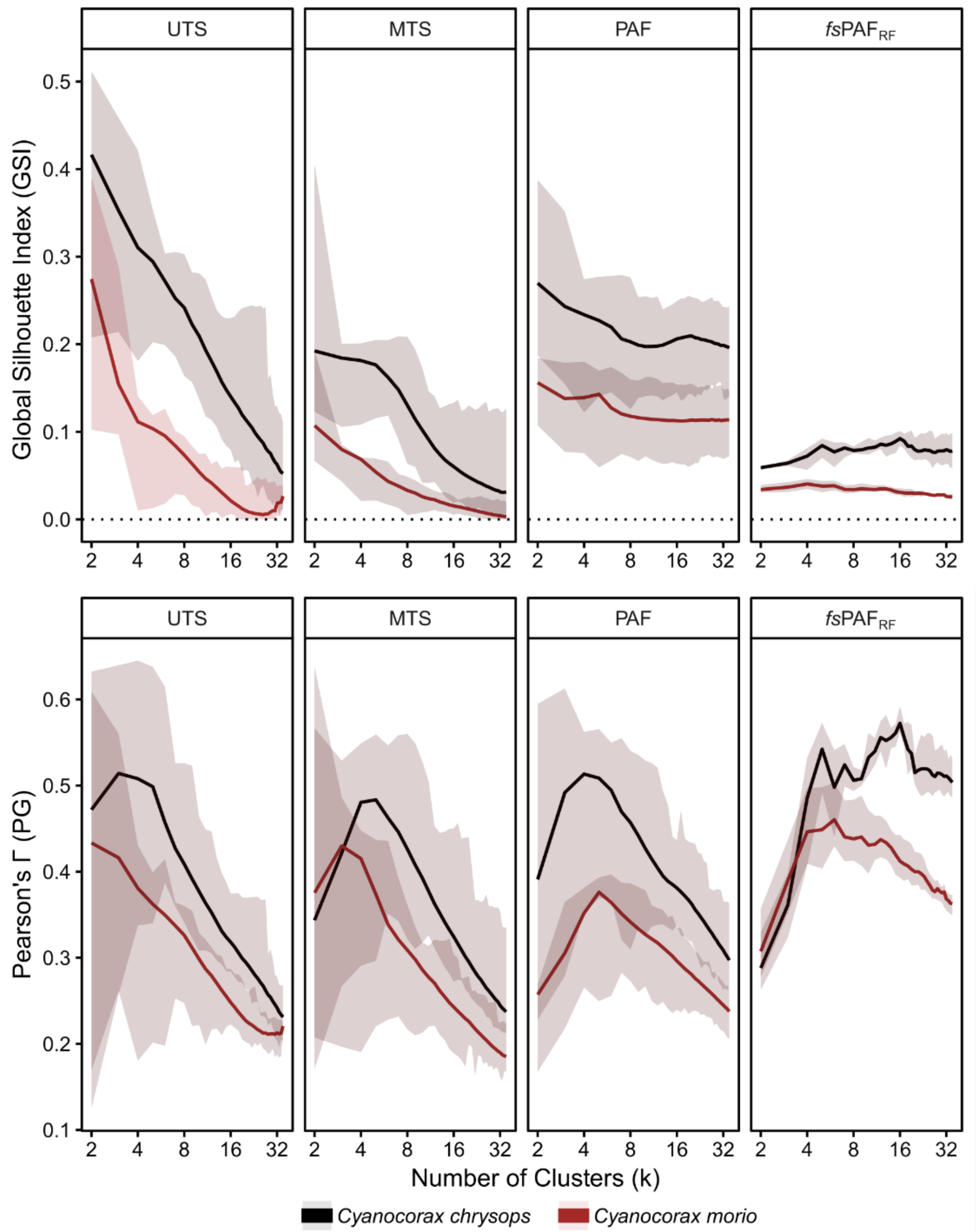


Figure S1. Highlight of the patterns in Global Silhouette Index (GSI) and Pearson's Γ (PG) between types of acoustic analysis used to represent vocalizations from *Cyanocorax chrysops* (black) and *Cyanocorax morio* (brown). Types of acoustic analysis are univariate time-series (UTS), multivariate time-series (MTS), and predefined acoustic features (PAF). In both metrics, larger values correspond to higher clustering signal. The feature space *fs*PAF$_{RF}$ was represented separately from other PAF because of its contrasting GSI and PG. The k-sweeps include 2 to 35 clusters, with k graphically represented in log scale. Shaded bands represent maximum and minimum values among clustering algorithms (except k-means and k-means++), while lines represent their averages.

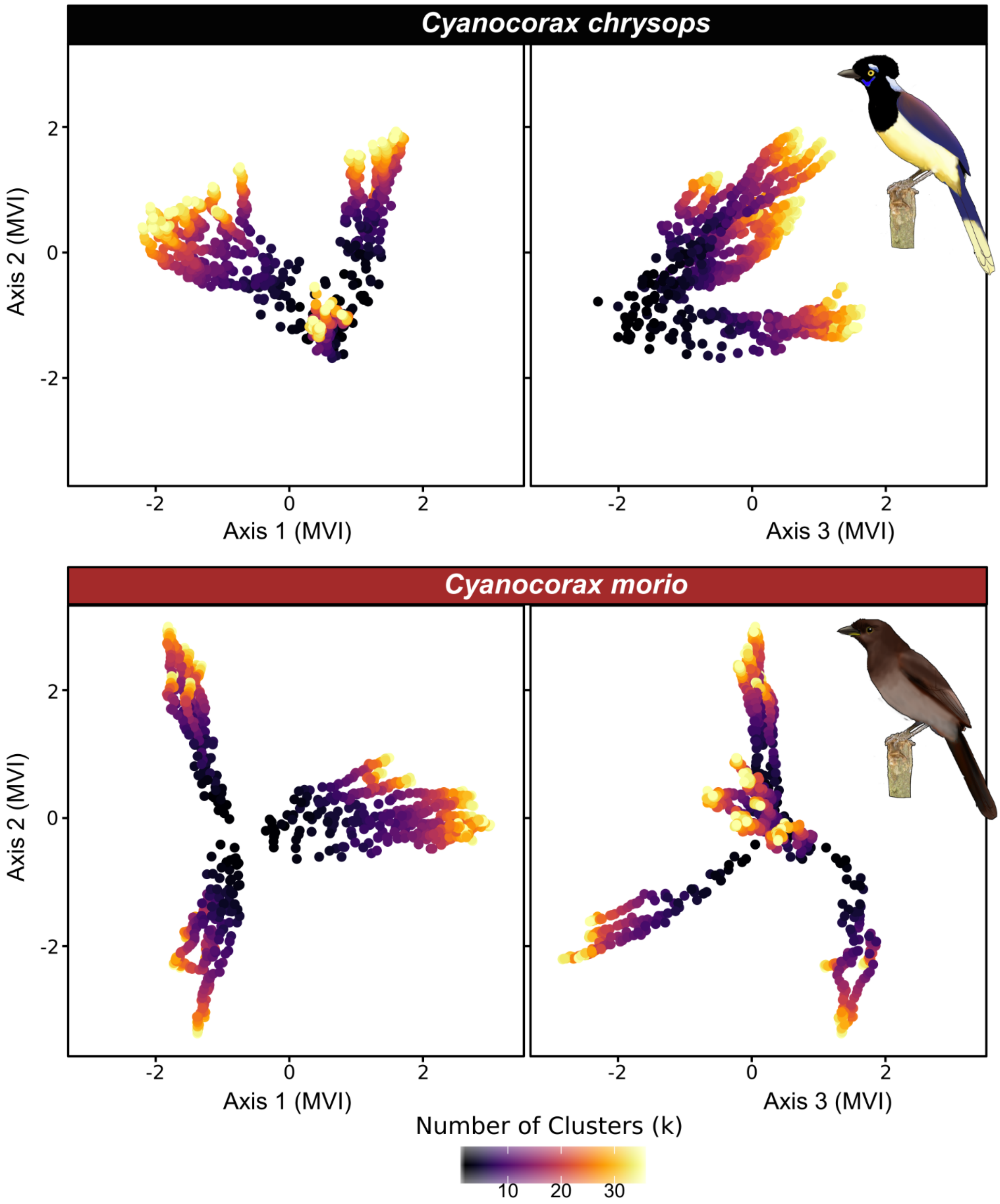


Figure S2. Ordination of pairwise Meilă's Variation of Information (MVI; Meilă, 2007) between labels obtained for *Cyanocorax chrysops* and *Cyanocorax morio* vocalizations from the tested clustering configurations. Different colours represent different numbers of clusters in the interval 2 ≤ k ≤ 35 to highlight how configurations diverge with increasing k, while relative variation is stable. Scales are the same in all panels.

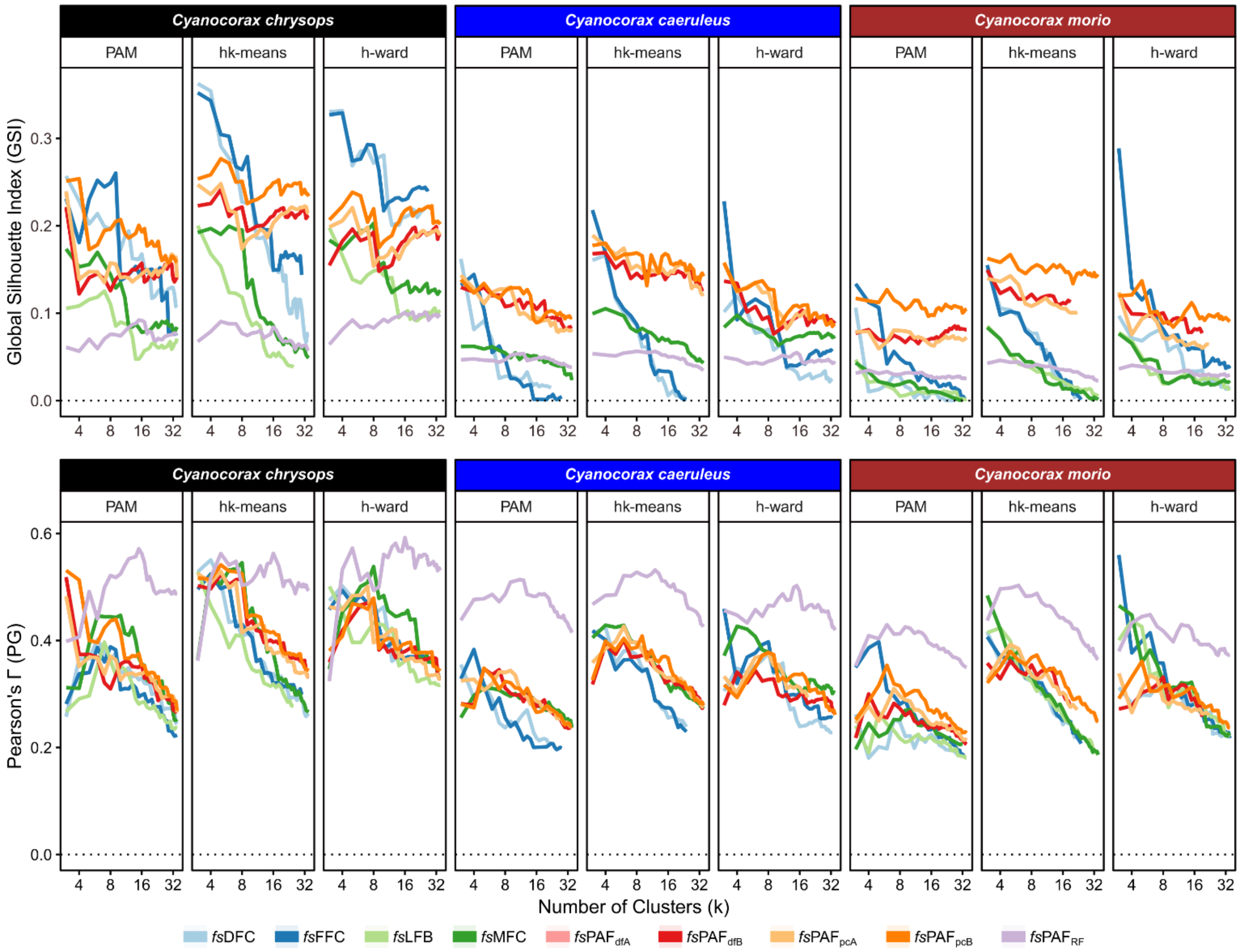


Figure S3. Global Silhouette Index (GSI) and Pearson's Γ (PG) for contextualization of clustering signal of data from *Cyanocorax coeruleus,* with data from *Cyanocorax chrysops* and *Cyanocorax morio* (already provided in Figure 1). The tested configurations were based in k-sweeps from 2 to 35 clusters, represented in log scale, different acoustic features are represented by different colours, and each facet a clustering algorithm (except k-means and k-means++). In both metrics, larger values correspond to higher clustering signal. Breaks in the k-sweeps represent failed configurations, such as negative silhouettes or those which generated clusters with only 5 samples or less. The feature spaces are: dominant frequency contours (*fs*DFC), fundamental frequency contours (*fs*FFC), linear frequency bins (*fs*LFB), mel-cepstral coefficients (*fs*MFC), predefined acoustic features (*fs*PAF$_{A}$), predefined acoustic features without variables with VIF superior to 10 (*fs*PAF$_{B}$), principal components of *fs*PAF$_{A}$ (*fs*PAF$_{pcA}$), principal components of *fs*PAF$_{B}$ (*fs*PAF$_{pcB}$), and random forest dissimilarity from the same predefined acoustic features (*fs*PAF$_{RF}$). Warm colours represent predefined acoustic features (PAF), tones of blue represent univariate time-series and tones of green represent multivariate time-series.

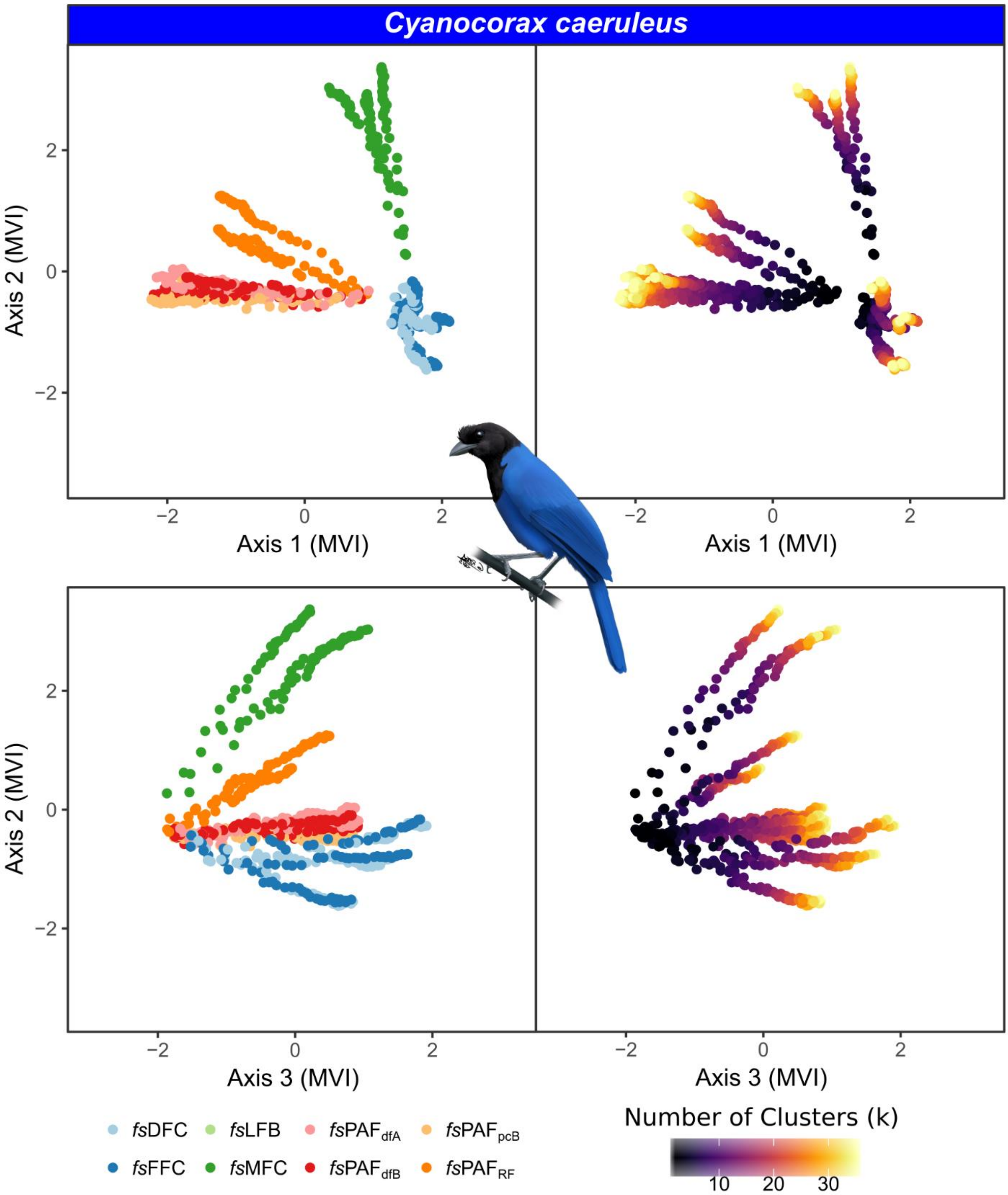
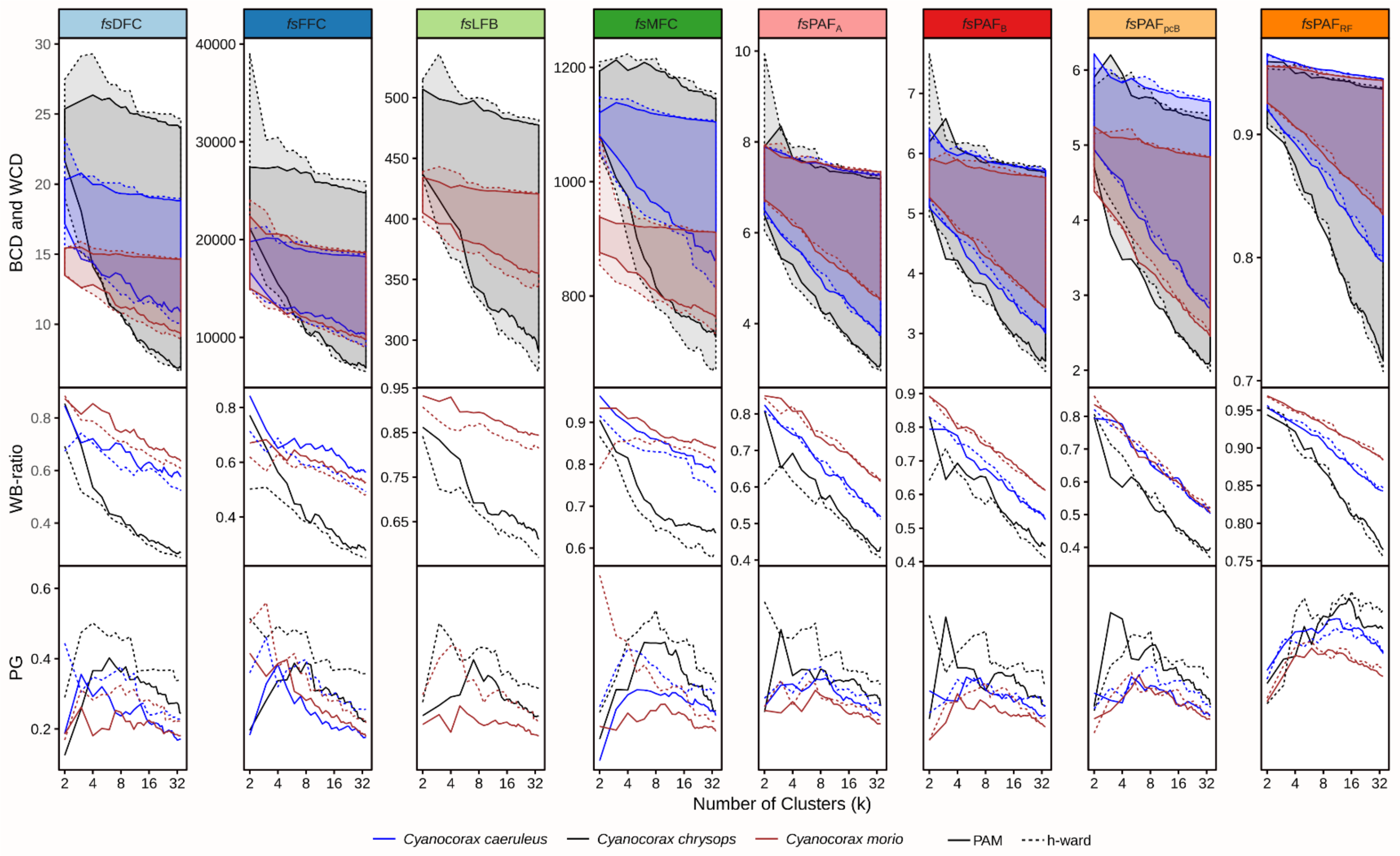


Figure S4. Ordination of pairwise Meilă's Variation of Information (MVI; Meilă, 2007) contextualization of data from *Cyanocorax coeruleus,* with data from *Cyanocorax chrysops* and *Cyanocorax morio* (already provided in Figure 2). The two columns are representations of the same MVI axes with two different colour codes. In the first column, colours represent the following feature spaces: dominant frequency contours (*fs*DFC), fundamental frequency contours (*fs*FFC), linear frequency bins (*fs*LFB), mel-cepstral coefficients (*fs*MFC), predefined acoustic features ($fsPAF_A$), predefined acoustic features without variables with VIF superior to 10 ($fsPAF_B$), principal components of $fsPAF_A$ ($fsPAF_{pcA}$), principal components of $fsPAF_B$ ($fsPAF_{pcB}$), and random forest dissimilarity from the same predefined acoustic features ($fsPAF_{RF}$). Warm colours represent predefined acoustic features (PAF), tones of blue represent univariate time-series and green represent multivariate time-series. Scales are the same in all panels. In the second column, colours represent different numbers of clusters in the interval $2 \le k \le 35$ to highlight how configurations diverge with increasing k, while relative variation is stable. Scales are the same in all panels.


fsDFC
fsFFC
fsLFB
fsMFC
fsPAF_A
fsPAF_B
fsPAF_pcB
fsPAF_RF
BCD and WCD
WB-ratio
PG
Number of Clusters (k)
Cyanocorax caeruleus
Cyanocorax chrysops
Cyanocorax morio
PAM
h-ward

876

Figure S5. Clustering features of *Cyanocorax coeruleus* (blue), *Cyanocorax chrysops* (black) and *Cyanocorax morio* (brown), produced by algorithms PAM (solid lines) and h-ward (dashed lines), to cluster vocalizations in k-sweeps from 2 to 35 clusters. The feature spaces are: dominant frequency contours (*fs*DFC), fundamental frequency contours (*fs*FFC), linear frequency bins (*fs*LFB), mel-cepstral coefficients (*fs*MFC), predefined acoustic features (*fs*PAF$_{A}$), predefined acoustic features without variables with VIF superior to 10 (*fs*PAF$_{B}$), principal components of *fs*PAF$_{B}$ (*fs*PAF$_{pcB}$), and random forest dissimilarity from the same predefined acoustic features (*fs*PAF$_{RF}$). To illustrate interspecific differences, dissimilarities between clusters (BCD) and within clusters (WCD) are represented as the upper and lower limits of shaded bands. The ratio between those clustering features is represented below as WB-ratio, along with clustering diagnostics Global Silhouette index (GSI) and Pearson's Γ (PG) included as visual reference. With a dotted line we represented $k \geq 8$ as the interval of interest, providing stable results for interspecific comparisons. The axis k is represented in log scale.

CAPÍTULO 2

Geographic variation in vocalizations of the Azure Jay (*Cyanocorax coeruleus*) reveals divergence between their eastern and western populations

Artigo a ser submetido ao periódico Journal of Avian Biology – http://www.avianbiology.org/authors/author-guidelines

Geographic variation in vocalizations of the Azure Jay (*Cyanocorax coeruleus*) reveals divergence between their eastern and western populations

Gabriel L. M. Rosa[1, 2] & Luiz dos Anjos[1, 3]

[1]Departamento de Biologia Animal e Vegetal, Universidade Estadual de Londrina, Rod. Celso Garcia Cid, 10.011, Brazil
[2]Orcid ID: 0000-0001-5332-9713
[3]Orcid ID: 0000-0002-8680-2375

Corresponding author: Gabriel L. M. Rosa, Departamento de Biologia Animal e Vegetal, Universidade Estadual de Londrina, Rod. Celso Garcia Cid, 10.011, Brazil. E-mail: g.rosa.bio@gmail.com

ABSTRACT

Geographic variation in phenotypes can be shaped by many biological processes, and in many cases could provide insight into previously unknown biodiversity. The unstable historical distribution of the Brazilian Atlantic Forest, and its forest types, is known to be shaped by phylogeographical patterns of many forest-dependent organisms. The threatened Azure Jay (*Cyanocorax coeruleus*), the most austral corvid in the New World, is endemic and one of the flag species of the southern Atlantic Forest. Divided by the Serra do Mar, a potential topographical barrier, their geographic distribution can be separated in two ecologically distinct regions: eastern and western. In this study, we explored geographic variation in vocalizations from this species to test if phenotypic divergence exists between populations of this species. Multiple statistical approaches confirmed increases in differences between vocalizations to be related to increases in geographic distance, a strong evidence of clinal variation shared by eastern and western populations of the Azure Jay. However, we provide the first evidence of divergence independent of the distance between them. Also, the effect of geographic distance was partially disrupted at Serra do Mar, indicating the presence of clinal variation exclusive to each population. A thorough evaluation of those patterns with machine learning methods revealed eastern-like features to be more spatially restrict than we anticipated in our hypothesis. We discuss how our results could relate to past distributions of the Atlantic Forest and its implications for the conservation of the Azure Jay. If future studies including genetic data confirm our results, it would emphasise the importance of sound collections and Bioacoustics tools to assess cryptic diversity.

Keywords: Bioacoustics, call, dialect, phenotypic variation, Atlantic Forest

INTRODUCTION

Geographic variation in phenotypes has often been used as a first line of evidence for previously unknown phylogeographic patterns (Avise et al. 2016, Zamudio et al. 2016), i.e. the spatial distribution of lineages within a species. The geographic variation in bird songs and calls are particularly well-studied examples (Slabbekoorn and Smith 2002, Podos and Warren 2007). Evidence of correlation with common inheritance can be found even in species that are able of learning and reshaping their vocalizations through life, such as cockatoos, parrots (Keighley et al. 2017, 2019, Wright and Dahlin 2017), and possibly corvids (Laiolo et al. 2001, Kryukov et al. 2017). Although the mechanisms that generate novelties in vocalizations and their inheritance, such as drift and selection, are sometimes difficult to assess (Slabbekoorn and Smith 2002), their spatial structure is not. When dispersal and inheritance are mainly limited by distance, phenotypical variation can form a continuous gradient, referred to as cline (Podos and Warren 2007). There are also cases in which isolation, of diverse causes and intensities, could significantly limit gene flow and social exchange, could produce at shorter scales and, divergence between clearly delimited units can also be possible (Podos and Warren 2007). When phenotypic variation of bird vocalizations is structured in a mosaic of such units, those units can be referred as to dialects (Marler and Tamura 1962).

Isolation in montane forests seems to be linked with the diversification of New World Jays (NWJ) of the genera *Aphelocoma* (Bhagabati et al. 2004, McCormack et al. 2008, 2010, Venkatraman et al. 2018) and *Cyanolyca* (Bonaccorso 2009). Bonaccorso (2009) and Venkatraman et al. (2018) explain this pattern as a possible consequence of limited dispersal. Venkatraman et al. (2018) diversification patterns of *Aphelocoma unicolor* resemble those of reptiles and amphibians from the same ranges. Limited dispersal linked to cooperative breeding seems to be a common feature among the NWJ (Brown 1974) and is well documented in other NWJ, such as *Calocitta formosa* (Langen 1996, Berg 2005) and *Cyanocorax morio* (Hale et al. 2003, Williams 2004, Williams and Rabenold 2005, Williams and Hale 2006). Indeed, NWJ have complex repertoires, but for most species, especially those from the genus *Cyanocorax,* several aspects of their life histories, behaviour, demography and phylogeographic patterns have yet to be studied properly (Rosa et al. 2016).

The most austral species, and maybe the most studied NWJ in South America, is *Cyanocorax coeruleus* (Vieillot, 1818). Its geographic range overlaps with the southern portion of the Atlantic Forest, which is separated by Serra do Mar in two ecologically distinct ecoregions (Veloso et al. 1991; Olson et al. 2001). Its eastern range, mostly near sea level, has a warm and humid climate in which dominant vegetation is the dense rain forest, also referred to as Serra do Mar ecoregion (Ribeiro et al. 2009). In the western range, higher elevation and colder climate favour mixed rain forest, also referred to as Araucaria ecoregion (Ribeiro et al. 2009), because of the dominance of *Araucaria angustifolia* (Bertol.) Kuntze, which is gradually fragmented and replaced by Campo grasslands at higher latitudes. Seasonal availability of *A. angustifolia* seeds is apparently a key resource for the survival of *C. coeruleus* during the colder winter in their western range (Boesing and Anjos 2011). Due to broad habitat loss and intense fragmentation, *A. angustifolia* is critically endangered (Thomas 2013). Therefore, *C. coeruleus* is considered near threatened mainly due to habitat loss (BirdLife International 2017a) and is considerably rare in Argentina and Paraguay, their westernmost historical range were habitat remnants are scarce (Argentini et al. 2018). Given the likely limitations for dispersal through Serra do Mar, divergence between populations found in the eastern and western ranges is a possibility, which will have important implications for the conservation of this threatened species. This hypothesis is yet to be tested. Different morphs involving different tones of blue are mentioned in the literature (Madge and Burn 1994, Ridgely and Tudor 2009, Brady 2010), which is an indicative that this species could be spatially structured. No aspect of intraspecific variation in their vocalizations has ever been investigated beyond the description of the species repertoire (Anjos and Vielliard 1993).

In the present study, we make the first assessment of intraspecific variation in the vocalizations of *C. coeruleus*, and investigate if the acoustic structure of their most common vocalizations (social-call *sensu* Anjos and Vielliard 1993) is geographically variable. We expect to find an increase in vocal differences with an increase of geographic distance, which will confirm a clinal variation. But our main objective is to test if the topographic and ecoregion divide represented by Serra do Mar could explain the divergence in acoustic features among the calls from the eastern and western populations of *C. coeruleus*. In order to thoroughly evaluate our results, we reassessed our hypothesis about boundaries between populations with another approach. This approach consisted in testing how well the population of origin can be predicted using only vocalization features. Although genetic data is not yet available, we discuss our hypothesis about how past distributions of the Atlantic Forest could be related to our findings and implications of these for the conservation of *C. coeruleus*. To the best of our knowledge, this is the first investigation of geographic variation in vocalizations of a NWJ after Bhagabati et al. (2004) mentioned potential variation in *Aphelocoma*, and Brunetta and Anjos (2010) investigated geographic variation in social-calls from *Cyanocorax chrysops*.

MATERIALS AND METHODS

Studied species

Cyanocorax coeruleus is the most austral member of the lineage of New World Jays, one of the lineages of the family Corvidae which expanded and diversified in the American Continent during the Miocene-Pliocene transition (Sumudu et al. 2017). Most of their geographic distribution is restricted to Southern Brazil, with sporadic records reaching the southern Paraguay border and Missiones, in North-eastern Argentina. Currently confirmed by Argentini et al. (2018), their occurrence in Paraguay was previously questioned because several reports could actually refer to *Cyanocorax cyanomelas*, given the morphological resemblance of both species. *Cyanocorax coeruleus* require evergreen forests as habitat, and throughout most of their range its reproduction seems to be conditioned to the presence of *A. angustifolia* (Boesing and Anjos 2011). Ridgely and Tudor (2009) mention this species to be abundant in coastal Restinga vegetation. Mostly because of habitat loss, *C. coeruleus* is considered near threatened (BirdLife International 2017). According to Anjos and Shibatta (2010), *C. coeruleus* can sustain flights of up to 1km between forest fragments. Their colour is among the simplest found for NWJ, presenting most of the body dark blue, with head, nape, beak, breast and feet having a black colour. According to Madge and Burn (1994) their tone of blue varies individually between 'cobalt-blue', 'violet-blue' to 'greenish-blue'. Ridgely and Tudor (2009) mention the 'greenish-blue' tone as a scarce morph, however, biological explanations for this variation still lack empirical evidence. Observations about their reproduction made by Boesing and Anjos (2012) confirm that the species breed cooperatively, a widespread behaviour in NWJ (Brown 1974). The species' vocal repertoire is composed of 14 basic call types and intermediate forms between them (Anjos and Vielliard 1993). There is no sign of syntax in any of their vocalizations. Their most common vocalizations are the social-calls, which are composed by a single acoustic signal with clear harmonic structure, with a short upward and long downward frequency modulations. However, the acoustic structure of these vocalizations types can be variable in many different aspects, including duration, overall frequency, harmonic structure and presence of harsh sections of variable duration at their beginning or end, which has not been properly tested.

Sound processing and sample selection

Social-calls from *C. coeruleus* were sampled from recordings available at the Macaulay Library of Nature Sounds (Cornell Lab of Ornithology), WikiAves, and Xeno-Canto, and from recordings made in the field by GLMR and LDA (metadata available in Table S1 of Supp. Matt.). Most recordings made by LDA are the same analysed in the description of this species repertoire (Anjos and Vielliard 1993). We excluded from our analyses all recordings where geographic coordinates were absent and could not be approximated from the available metadata. We standardized the recordings to high-resolution audio (WAV; 44.1kHz; 24-bit), and within each recording, we selected calls that did not overlap any other signal and the energy was sufficiently above the background noise for an adequate acoustic analysis. Except for this last step, performed in Audacity *v*2.3.0 (Audacity Team 2019), with spectrogram parameters set for clear visualization of the start and end of each signal (Figure 1A; window = Hann; frequency scale = Mel; fft size = 512; overlap = 90%), the procedure from this point on was entirely performed in the R software environment (R Core Team 2019). Without including silence in-between calls, produced individual files for each call, further referred as to samples (Figure 1B). All energy considerably out of the known hearing range of corvids (300-6000Hz; Jensen and Klokker 2006) was attenuated to match the background.

Sound representation

We chose Mel-cepstral Coefficients (MFCC; package ‘tuneR’ *v*1.3.2; Ligges 2017) to represent the acoustic features of the samples in sound analyses and comparisons. Although compressed, the MFCC (Figure 1C) allows robust and objective comparisons between signals with complex energy and frequency modulation patterns (Ranjard and Ross 2008; Ranjard et al. 2010). The 12 cepstral coefficients of the MFCCs were calculated from modified Mel-scale cepstra (300-6000Hz; window = Hann; wl = 5ms), with pre-emphasis filter set to 0.97 and liftering coefficients set to 22, so differences in magnitude between cepstra were smaller (Young et al. 2006, Krull et al. 2012). For an in-depth exploratory analysis of our results, we also represent samples by a multivariate set of 27 predefined acoustic features (PAF; function 'specan' from the R package 'warbleR' *v*1.1.12; Araya-Salas and Smith-Vidaurre 2017; see Table S2 in Supp. Matt.). All PAF were measured using the same spectrogram parameters (Figure 1D; window = Hann; f = 44100; wl = 256; ovlp = 90%). In this representation, components can be intuitively interpreted and linked to biological patterns.

Data preparation

As the number of samples was not balanced among different recordings, we used a modified version of the strategy used by Krull et al. (2012) to obtain a single instance per recording. The final product of this strategy is an average MFCC representation, built through a recursive alignment with the Dynamic Time Warping algorithm (DTW; package 'dtw' *v*1.18-1; Giorgino 2009; and 'dtwclust' v5.1.0; for more details see Sarda-Espinosa 2017). This strategy was performed in all recordings containing at least two samples. In this method, for *N* different MFCC matrices, each representing a signal, in a list *L*:

(a) *If* there is a single element in L, *then* finish the procedure, *else* go to step (b);
(b) Calculate a pairwise distance matrix (*D*) between all elements of *L*;
(c) Identify the most distant elements in *D*, referred to as *a* and *b*, and compute their alignment (Figure 2) via warping path, as in Ranjard and Ross (2008);
(d) The averaged cepstra of all aligned windows in *a* and *b* are used to build a new MFCC matrix (*m*);
(e) The cepstra in *m* are interpolated, so the number of windows match the nearest integer to the average between the total number of windows of *a* and *b*;
(f) Remove *a* and *b*, add *m* into *L*, and recalculate *D*;
(g) *If* there is a single element in L, *then* go back to step (b), *otherwise* finish the procedure.

This procedure will iterate until a single dynamically aligned MFCC instance remains among those from each recording. To represent variation between recordings we obtained D^{voc}, a vocal distance matrix calculated using the DTW algorithm, in which normalized and unconstrained comparisons were made between the single MFCC we produced from each recording. Vocal distances represented in a Euclidean space, between the single MFCC instance produced from each recording. We chose this strategy to obtain D^{voc} because it minimizes the number of analytical decisions and avoids complex assumptions about which acoustic properties are more important.

To investigate the spatial structure of *C. coeruleus* social-calls, we produced a pairwise geographic distances matrix, D^{geo}, that matches D^{voc}. We calculated geographic distances in km using coordinates where the recordings were made as input, and accounting for the curvature of the earth using Haversine formula (Sinnot 1984). Because our objective was to test the potential divergences between samples from potentially different populations, we divided *C. coeruleus* geographic distribution in two ranges, eastern and western, and produced a representation matching D^{geo} and D^{voc}. To build it we assigned all samples collected below 500m of altitude (Fick and Hijmans 2017) in the eastern slope of Serra do Mar to the eastern population (E), and all remaining samples, collected west of the same boundary, to the western population (W). Though not perfect, this boundary closely matches the transition between the dense and mixed rainforests proposed by Olson et al. (2001). In all locations south from the southern limit of Serra do Mar (30ºS), where expressive topographic barriers are absent, we assigned all samples to the western population because similarly to the eastern range at lower latitudes, *C. coeruleus* presence is also dependent on *A. angustifolia*. Population assignments were used to produce a pairwise matrix with binary content representing population sharing, referred as to D^{pop}. In this case, the value 0 was assigned to all comparisons of samples made within the same range (E *vs*. E; W *vs*. W) and the value 1 to all comparisons made between different ranges (W *vs*. E).

Statistical Inference

The most essential question of our study was if there was an association between population assignments and vocal distances is greater than of that expected by chance. First, we tested if vocalizations from eastern and western populations have similar dispersion around their centroids, i.e. the mean positions in the ordination space produced from D_{voc} (PERMDISP2; function 'betadisper' in the package 'vegan' *v2.5-3*; Oksanen et al. 2018; *see* Anderson 2006). In other words, we tested if both populations have similar variance structures. Not rejecting the null hypothesis of similar dispersions is a prior requirement for the next step, in which we test if centroids from eastern and western populations are in significantly different locations (PERMANOVA; function 'adonis2' of the package 'vegan'). Rejecting the null hypothesis in this second step, means both populations have different locations in the ordination space produced from D^{voc}, or different averages.

As population assignment is spatially structured by definition, we verified the effects of geographic distance and its interaction with these assignments. We verified first if vocal distances in D^{voc} are spatially autocorrelated, i.e. positively correlated with corresponding geographic distances in D^{geo}. For such, we performed Mantel correlation test (Mantel 1967) between D^{voc} and D^{geo}, with 9999 randomizations of D^{voc}. In addition to this, we performed a partial Mantel Test of the correlation between the residuals of D^{voc} and D^{geo} after a linear correlation with D_{pop}. In other words, the partial test was performed to assess if vocalizations are spatially autocorrelated regardless of assignment to the same population or not.

As in some conditions the performance of Mantel correlation tests could be sub-optimal (*see* Diniz-Filho et al. 2013, *but also* Guillot and Rousset 2013, Legendre et al. 2015) we performed additional tests with an alternative approach (Lovette and Hochachka 2006, Yandell et al. 2018). This approach consists on testing the linear response of D^{voc} to D^{geo}, D^{pop}, and to the interaction of D^{geo} and D^{pop}. The regression intercept is interpreted as vocal distance in the immediate surroundings of the samples; D^{geo} coefficient as the ratio of increase in vocal distances related to increase in geographic distance; D^{pop} coefficient represents the amount of variation in vocal distances between populations that is not explained by geographic distance; and the interaction between D^{geo} and D^{pop} as the difference between the ratios of vocal distance increase in both populations. Given that pairwise comparisons are not independent instances, we compared the coefficients obtained with the original data to their empirical null distributions. To calculate those distributions, we performed 1000 randomizations of the response variable, D^{voc}, and fitted the original predictors and stored intercepts and coefficients for each iteration. In this approach, a predictor which coefficient outside the 95% confidence limits of its null distribution ($\alpha = 0.05$) has a non-random relationship with D^{voc}. A schematic representation of the statistical procedures described above is available as Figure S1.

As proposed by Yandell et al (2018), there are five partially competing hypotheses for predictions based in the relationship between D^{voc} with D^{geo} and D^{pop}, when the last discriminates comparisons ‘within populations’ from those ‘between populations’ (see Figure 3). If there is no effect of D^{geo}, i.e. no clinal variation, coefficients of comparisons within and between populations are not different from their null distributions. In this case, if the intercepts are also the same, variation in the data is not linked to linked to geographic distance or to population divergence (Figure 3A). But, if intercepts differ, the data supports population divergence, though it is not related to geographic distance (Figure 3B). In case coefficients of comparisons within and between populations are outside their null distributions, their effect is significant (not null), therefore the data support evidence of clinal variation. Though, in the presence of clinal variation, three outcomes are possible. In the case of geographic effect, coefficients and intercepts are the same, supporting the existence of spatially structured vocalizations in a single cline, but without population divergence (Figure 3C). In the case of non-interacting effect, coefficients are the same but intercepts differ, supporting population divergence and similar ratios of increase in vocal distance through geographic distances, regardless of the population (Figure 3D). At last, in the case of an interaction effect, coefficients and intercepts differ, supporting that populations diverge and their ratios of increase in vocal distance through geographic distances are different (Figure 3E). We expect to find agreement between Mantel tests and the alternative from the above which obtained support with our data.

Additionally, to aid in the interpretation of our results, we performed a direct inspection of the first two axes of an NMDS (Gower 1966) of D_{voc} in geographic space. Geographic coordinates of the points were rotated by 0.4 radians, so the geographic separation between the two populations is represented orthogonally to the first axis, the modified longitude. A local polynomial regression (LOESS; Cleveland et al. 1992) was fitted for each combination of NMDS axes and rotated coordinates. We used unit span to provide a balanced representation of local and general trends.

Class Prediction

Despite its robustness, the DTW algorithm produces comparisons, not descriptions. Therefore, we resorted to PAF for a more in-depth exploration of what aspects of the vocalizations are related to the patterns we found. We trained a Random Forest (package ‘randomForest’ v.4.6-14; Liaw and Wiener 2002, see also Breiman 2001, Keen *et al.* 2014) to predict population assignments of all samples, represented by their PAF. This algorithm performs supervised classification based in PAF from samples, including information about populations of origin (train; 20% of the instances), and predicts population labels for the remainder (test; 80% of the instances) by voting in the outcomes of multiple random decision trees (n = 1000; 4 variables at each split).

The most voted decision, given the posterior probability of belonging to either population, is taken as the final prediction for each instance. One of the main advantages of the Random Forest algorithm is that variable scale, highly heterogeneous in our set of PAF, does not affect the results, thus avoiding pre-processing steps such as standardization. From the classification results, we assessed how well PAF support population differences, and which acoustic features contributed more for its accuracy (Gini Index; see Liaw and Wiener 2002), and, most importantly, the posterior probabilities of population assignment for all classified samples, i.e. how these fit in what would be expected for each population. A schematic representation of the classification procedure described above is available in Figure S2.

RESULTS

We analysed 5398 social-call samples from *C. coeruleus*, in which vocal distances between averaged MFCC (D^{voc}; n = 166) contained evidenced geographic variation. Our hypothesis about divergence between eastern and western populations was confirmed (PERMANOVA; SS = 192.65; pseudo-F = 10.664; R^2 = 0.061; P < 0.001), with similar variance in both populations (PermDISP; F = 0.152; p = 0.704). Clinal variation was also present, as vocal distances and geographic distances were correlated (Pearson's r = 0.1485; Mantel test p = <0.001). This result is, at least in part, independent of the population identity (Pearson's r = 0.119; Partial Mantel test p = <0.001).

The independent effect of geographic distances is confirmed by positive and significant coefficients, while the independent effect of population sharing is confirmed by the prediction of different intercepts (Table 1). In turn, the interaction between those effects is also significant (Table 1), and could not be removed from the model without significant loss of explanatory power (SS = -289.03; F = 181.76; P < 0.001). Predictions showed lower intercepts and higher coefficients for comparisons sharing the same population, when compared to those from distinct populations (see Table 1). The effect of geographic distance over vocal distance was clearer within populations (Figure 4), matching the 'interaction effect' among competing hypothesis (Figure 1E). It implies clinal variation in the species' geographic distribution, which is partially disrupted at the boundaries between eastern and western populations. All coefficients in this model were outside the confidence intervals of their null distributions (Table 1). The inspection of directional trends in NMDS axes of D^{voc} in rotated coordinates revealed continuity between phenotypes of both populations. The most remarkable discontinuity can be found in directional trends of MDS2 along the rotated longitude (see Figure 5), in which the larger discontinuity between phenotype variation overlaps Serra do Mar. We interpreted such directional trends, including that of MDS2 in the rotated latitude, as clinal variation with varying levels of overlap.

The Random Forest classification of samples (n = 4317) based on their PAF had a low error rate (OOB error = 13.04%). However, classification accuracy was biased toward the western range (class error = 3.98%), in detriment of the eastern range (class error = 42.68%). This is likely a consequence of a high number of samples recorded South from 28º S inside the eastern range that were classified as western, or at least ambiguous (Figure 6; see also Figure S3). Therefore, within the eastern range, the spatial distribution of samples unambiguously classified as eastern-like is rather restricted to locations north-east of 28ºS 49ºW. Empirical assignments based on predictions made using only PAF revealed clearer gaps between ranges than our original hypothesis (Figure 7). Posterior probabilities are available in Table S1. The eight acoustic features with the highest contrasts between populations were mostly related to entropy (‘sfm’, ‘sp.ent’ and ‘entropy’), the median of the energy distribution in frequency (‘freq.median’), the skew of dominant frequency modulation (‘skew’), the instant in time of the first quartile of energy (‘time.Q25’), the mean dominant frequency (‘meandom’) and the frequency of the energy peak (‘peakf’).

DISCUSSION

According to our findings, increase in geographic distances between sampling locations explained the increase in acoustic distances between social-calls of *C. coeruleus*. However, this effect was smaller when comparisons were made within the eastern and western populations than between populations. The relevance of geographic distance could be an indication of limited dispersal, such is the case for documented in other closely related NWJ (Langen 1996, Hale et al. 2003, Williams 2004, Berg 2005, Williams and Rabenold 2005, Williams and Hale 2006). Similarly to Yandell et al. (2018), geographic distance and population sharing interact, which appears to be unique clinal variation within each population as well as for the whole species range. Because variation was large, even between samples from the same recording, the role of other biological factors not included in this study could be relevant. For this reason, we limited our interpretation to the patterns consistent at larger scales. A closer look at where in space those divergences are located revealed a more complex scenario than we anticipated.

A thorough inspection of PAF revealed eastern-like social-calls as less organized in their spectra (‘sfm’, ‘sp.ent’ and ‘entropy’), high pitched (‘freq.median’, ‘meander’), skewed in modulation towards their starts (‘skew’) and more gradual in their initial amplitude modulation (‘time.Q25’), relative to the western-like social-calls. Those features had the largest contributions for correct populations predictions of origin for social-calls. We also found that western-like features were also more variable (Figure 4), as expected because of the larger geographic range for western populations. A superficial cross-species assessment (pers. obs.), readily reveal the above features as the most obvious differences between social-calls of *C. coeruleus* and their closely related species (clade A in Bonaccorso et al. 2010). The most common calls of *Cyanocorax cyanomelas*, *Calocitta colliei* and *Calocitta formosa*, are harsh and apparently show higher entropy, while *Cyanocorax cristatellus* and *Cyanocorax violaceus* are intermediate, and *C. coeruleus* and *Cyanocorax morio*, closer to low entropy pure tones. We suggest that such acoustic features could be correlated to common ancestry, and, in turn, make vocalizations a reliable source of information about the species’ phylogeographic patterns. Nevertheless, biological explanations for acoustic variation in corvid vocalizations may also include the acoustic adaptation to different habitats (Nishimura et al. 1991, Laiolo and Rolando 2003), different levels of motivation (Yorzinski and Vehrencamp 2009), individual signatures (Kondo et al. 2010), and traits related to age and sex in diverse contexts (Hameed et al. 2014, Mates et al. 2014, Rosa et al. 2016). Possibly, lower frequencies of western-like social calls could be an allometric side-effect of a larger body mass of a population adapted to the colder climate of the western range, though the opposite is not necessarily true (Laiolo and Rolando 2004, Gillooly and Ophir 2010, Rodríguez et al. 2015, Torres et al. 2017).

Class predictions based in PAF motivated us to re-evaluate our initial hypothesis about the spatial distribution of western-like and eastern-like features. From posterior probabilities obtained, we detected several western-like social-calls recorded inside the eastern range. Dominance of eastern-like features were present in the entire eastern population, but were dominant only in its northern half (<28ºS). Even so, samples with western-like features or intermediate become increasingly common when approaching Serra do Mar. Sample reassignment following these predictions noticeably improved the gap in population comparisons regarding the features mentioned above. In this scenario, the range in which eastern-like social-calls are dominant is smaller than it was in our original hypothesis, as several western-like samples were found within the eastern range (Figure 3). A phylogeographic hypothesis that could explain our results would implicate divergence followed by multiple events of secondary contact, ultimately resulting in introgression of western phenotypes into the eastern population. This process could be a product of the past climate events, which are known to result in highly unstable forest cover in the southern portion of the Atlantic Forest since the Pleistocene (Behling and Negrelle 2001, Behling 2007, Behling and Pillar 2007, Carnaval and Moritz 2008, Carnaval et al. 2009, 2014, Thomé et al. 2010, Ledru et al. 2016). Palynological records dated tropical forest expansion South of -26ºS in the eastern range of Serra do Mar as a recent event, consolidating only after 6700 ^{14}C yr B.P. (Behling and Negrelle 2001). Continuity between forest covers of both ranges would have happened first in -29ºS, when grasslands were replaced by *A. angustifolia* dominated forests during the late Holocene (4320 ^{14}C yr B.P. to present, Behling and Pillar 2007). The dominance of grasslands in this first contact was unchanged since the Last Glacial Maximum (LGM) and through most of the Holocene (42,840 uncal. and 11,500 ^{14}C yr B.P.). Palynological records taken above 1000m at -26ºS indicate that Serra do Mar highlands were covered exclusively by grasslands since the last Glacial maximum (42,000-18,000 B.P.), with a following expansion of *A. angustifolia* to higher elevations only after 2000±500 B.P. (Behling 2007). Therefore, after the LGM, as the climate warmed, grasslands retracted to higher altitudes and were replaced by forests, which eventually connected the valleys across Serra do Mar. Then, secondary contact would have been established at multiple places, as isolation by Serra do Mar weakened. Phylogeographic patterns of forest-dependent birds and anurans can be valuable contributions to the knowledge about the historical distribution of the Atlantic Forest (Cabanne et al. 2007, Carnaval et al. 2009, Thomé et al. 2010). *Cyanocorax coeruleus* can be such a species, if time estimations of genetic divergence confirm this. However, estimates considering an assemblage of *Cyanocorax* (*C. coeruleus* not included), Sumudu and Peterson (2017) dated the divergence to present standards at late Pleistocene, far before the LGM.

If phenotypic divergence we found is correlated to genetic diversity of the study species, it would be necessary to reassess its taxonomic status, as well as its threat level and conservation strategies, which would be different between eastern and western ranges. The eastern range of *C. coeruleus* overlaps with the biological sub-region of Serra do Mar (*sensu* Ribeiro et al. 2009), which has an approximate area of 11.5 million hectares in which 32.2.% of the original forest cover remains, and only 25.2% of those are inside protected areas (Ribeiro et al. 2009). The biological sub-region of Araucaria (*sensu* Ribeiro et al. 2009), which overlaps the western range of *C. coeruleus*, has an approximate area of 25.3 million hectares, of which 12.6% of the original forest cover remains, and only 0.65% are inside protected areas (Ribeiro et al. 2009). Total protected areas in the Araucaria sub-region are 13.6% of the equivalent in the Serra do Mar sub-region, despite having an area more than twice as large. Five of the largest protected remnants in the entire Atlantic Forest (~100,000 ha) are in Serra do Mar region, while only two areas in the Araucaria region are larger than 20,000 ha (Ribeiro et al. 2009). Beyond this discrepancy in the proportion of protected areas between both ranges, logging and livestock farming and ranching are considerable threats to *A. angustifolia* in the Western range. Further, recruitment, regeneration, and population structure could also be affected by annual harvest of their seeds for human consumption (Thomas 2013, see also Souza 2007, Paludo et al. 2010, 2011). This scenario raises significantly the threat on the western population viability, which would be far more threatened by habitat loss and resulting fragmentation. Although conservation status and protection of forests are higher in the eastern range, the new concern for the population of *C. coeruleus* in this range would be the vulnerability to local impacts, given the high level of endemism. Therefore, there is the need to increase the number and size of protected areas of Araucaria forest for the maintenance of the western population of *C. coeruleus*.

Our limitation in the present study is the lack of data about this species' acoustic profile regarding the effect of sex, age and social roles. Ideally, information about the role of social learning would also be valuable to account for inheritance which is not correlated with genetics. This is a critical aspect to understand vocal communication in social organisms with highly developed cognitive abilities such as parrots (Dahlin et al. 2014, Wright and Dahlin 2017) and potentially NWJ (Rosa et al. 2016). Laiolo et al. (2001a, b) highlight how the complexity of corvid repertoires could be a reason why such studies are scarce. Maybe the most important venue for further research is the validation of the pattern we found with the use of genetics. If eastern and western populations are indeed genetically different, vocal phenotypes would then be reliable indicators of phylogeographic patterns to be explored in a broader phylogenetic scale. There is also the possibility of congruence between *C. coeruleus* past and the unstable history of forests in the extreme south of the Atlantic Forest. Thus, sampling geographic gaps and increasing spatial resolution of data around Serra do Mar would could then reveal a detailed picture of this species history. Beyond a window to the past, consistent evidence of divergence would require a reassessment of *C. coeruleus* conservation status along with an urgent call for conservation of critically endangered *A. angustifolia*.

ACKNOWLEDGEMENTS

We thank Macaulay Library, WikiAves, and Xeno-canto, for making available most recordings used in this study. Carlos B. de Araújo, David. D. Yandell, Emma Greig, Marcos R. Lima and Ricardo P. Adelino for valuable advice about the data analyses.

FUNDING

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

CONFLICTS OF INTEREST

The author declares no conflicts of interest.

References

Anderson, M. J. 2006. Distance-based tests for homogeneity of multivariate dispersions. – Biometrics 62: 245–253.

Anjos, L. and Shibatta, O. A. 2010. On the habitat occupancy of some neotropical jays in South America, genus *Cyanocorax*. – Semin. Ciências Biológicas e da Saúde 31: 153–158.

Araya-Salas, M. and Smith-Vidaurre, G. 2017. warbleR: an R package to streamline analysis of animal acoustic signals. – Methods Ecol Evol. 8: 184–191.

Argentini, N., Smith, P., Rodríguez, O., Castillo, H. and Ríos, S. D. 2018. The Azure Jay *Cyanocorax caeruleus* (Aves:Corvidae) in Paraguay, with restriction of the type locality. – Rev. Bras. Ornitol. 26: 210–213.

Audacity Team 2019. Audacity(R): Free Audio Editor and Recorder. Version 2.3.1.

Avise, J. C., Bowen, B. W. and Ayala, F. J. 2016. In the light of evolution X: Comparative phylogeography. – Proc. Natl. Acad. Sci. 113: 7957–7961.

Berg, E. C. 2005. Parentage and reproductive success in the white-throated magpie-jay, *Calocitta formosa*, a cooperative breeder with female helpers. – Anim. Behav. 70: 375–385.

Bhagabati, N. K., Brown, J. L. and Bowen, B. S. 2004. Geographic variation in Mexican jays (*Aphelocoma ultramarina*): Local differentiation, polyphyly or hybridization? – Mol. Ecol. 13: 2721–2734.

BirdLife International 2017. *Cyanocorax coeruleus* (amended version of 2016 assessment). – The IUCN Red List of Threatened Species 2017.

Boesing, A. L. and Anjos, L. Dos 2012. The Azure Jay may reproduce in plantations of *Araucaria angustifolia* in southern Brazil. – Bird Conserv. Int. 22: 205–212.

Bonaccorso, E. 2009. Historical biogeography and speciation in the Neotropical highlands: Molecular phylogenetics of the jay genus *Cyanolyca*. – Mol. Phylogenet. Evol. 50: 618–632.

Bonaccorso, E., Peterson, A. T., Navarro-Sigüenza, A. G. and Fleischer, R. C. 2010. Molecular systematics and evolution of the *Cyanocorax* jays. – Mol. Phylogenet. Evol. 54: 897–909.

Brady, S. 2010. Azure Jay (*Cyanocorax caeruleus*), version 1.0. – In: T. S. Schulenberg (ed.), Neotropical Birds Online. Cornell Lab of Ornithology, Ithaca.

Breiman, L. 2001. Random forests. – Mach. Learn. 45: 5–32.

Brunetta, B. and Dos Anjos, L. 2010. Variações na distribuição espacial e no grito social da gralha-picaça (*Cyanocorax chrysops*) na Mata Atlântica, Sul do Brasil. – Ornitol. Neotrop. 21: 203–213.

Cleveland, W. S., Grosse, E. and Shyu, W. M. 1992. Chapter 8: Local regression models. – In: Chambers, J. M. and Hastie, T. J. (eds.). Statistical Models in S. Wadsworth & Brooks/Cole.

Dahlin, C. R., Young, A. M., Cordier, B., Mundry, R. and Wright, T. F. 2014. A test of multiple hypotheses for the function of call sharing in female budgerigars, *Melopsittacus undulatus*. – Behav. Ecol. Sociobiol. 68: 145–161.

Diniz-Filho, J. A. F., Soares, T. N., Lima, J. S., Dobrovolski, R., Landeiro, V. L., Telles, M. P. de C., Rangel, T. F. and Bini, L. M. 2013. Mantel test in population genetics. – Genet. Mol. Biol. 36: 475–485.

Fick, S. E. and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. – Int. J. Climatol. 37: 4302–4315.

Giorgino, T. 2009. Computing and Visualizing Dynamic Time Warping Alignments in R : The dtw Package . – J. Stat. Softw. 31: 1–24.

Gower, J. C. 1966. Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis. – Biometrika 53: 325.

Guillot, G. and Rousset, F. 2013. Dismantling the Mantel tests. – Methods Ecol. Evol. 4: 336–344.

Hale, A. M., Williams, D. A. and Rabenold, K. N. 2003. Territoriality and Neighbor Assessment in Brown Jays (*Cyanocorax morio*) in Costa Rica. – Auk 120: 446–456.

Jensen, K. K. and Klokker, S. 2006. Hearing sensitivity and critical ratios of hooded crows (*Corvus corone cornix*). – J. Acoust. Soc. Am. 119: 1269.

Keen, S., Ross, J. C., Griffiths, E. T., Lanzone, M. and Farnsworth, A. 2014. A comparison of similarity-based approaches in the classification of flight calls of four species of North American wood-warblers (Parulidae). – Ecol. Inform. 21: 25–33.

Keighley, M. V., Langmore, N. E., Zdenek, C. N. and Heinsohn, R. 2017. Geographic variation in the vocalizations of Australian palm cockatoos (*Probosciger aterrimus*). – Bioacoustics 26: 91–108.

Keighley, M. V., Heinsohn, R., Langmore, N. E., Murphy, S. A. and Peñalba, J. V. 2019. Genomic population structure aligns with vocal dialects in Palm Cockatoos

(*Probosciger aterrimus*); evidence for refugial late-Quaternary distribution? – Emu 119: 24–37.

Krull, C. R., Ranjard, L., Landers, T. J., Ismar, S. M. H., Matthews, J. L. and Hauber, M. E. 2012. Analyses of sex and individual differences in vocalizations of Australasian gannets using a dynamic time warping algorithm. – J. Acoust. Soc. Am. 132: 1189–1198.

Kryukov, A. P., Spiridonova, L. N., Mori, S., Arkhipov, V. Y., Red'kin, Y. A., Goroshko, O. A., Lobkov, E. G. and Haring, E. 2017. Deep Phylogeographic Breaks in Magpie *Pica pica* Across the Holarctic: Concordance with Bioacoustics and Phenotypes. – Zoolog. Sci. 34: 185–200.

Laiolo, P. and Rolando, A. 2003. The evolution of vocalisations in the genus Corvus: Effects of phylogeny, morphology and habitat. – Evol. Ecol. 17: 111–123.

Laiolo, P., Rolando, A., Delestrade, A. and de Sanctis, A. 2006. Geographical Variation in the Calls of the Choughs. – Condor 103: 287.

Laiolo, P., Rolando, A., Delestrade, A. and Sanctis, A. De 2001. Geographic diversification in the call repertoire of the genus *Pyrrhocorax* (Aves, Corvidae). – Can. J. Zool. 79: 1568–1576.

Langen, T. A. 1996. Skill acquisition and the timing of natal dispersal in the white-throated magpie-jay, *Calocitta formosa*. – Anim. Behav. 51: 575–588.

Ledru, M. P., Montade, V., Blanchard, G. and Hély, C. 2016. Long-term Spatial Changes in the Distribution of the Brazilian Atlantic Forest. – Biotropica 48: 159–169.

Legendre, P., Fortin, M. J. and Borcard, D. 2015. Should the Mantel test be used in spatial analysis? – Methods Ecol. Evol. 6: 1239–1247.

Liaw, A. and Wiener, M. 2002. Classification and Regression by randomForest. – R News 2*:* 18–22.

Ligges, U., Krey, S., Mersmann, O., and Schnackenberg S. 2018. tuneR: Analysis of Music and Speech. R package version 1.3.2.

Lovette, I. J. and Hochachka, W. M. 2006. Continent-wide surveys demonstrate simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. – Ecology 87: S14.

Madge, S. and Burn, H. 1994. Crows and jays: a guide to the crows, jays and magpies of the world. – C Helm, p. 85.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. – Cancer Res. 27: 209–220.

Marler, P. and Tamura, M. 1962. Song “Dialects” in Three Populations of White-Crowned Sparrows. – Condor 64: 368–377.

McCormack, J. E., Peterson, A. T., Bonaccorso, E. and Smith, T. B. 2008. Speciation in the highlands of Mexico: Genetic and phenotypic divergence in the Mexican jay (*Aphelocoma ultramarina*). – Mol. Ecol. 17: 2505–2521.

Nishimura, K., Nakamura, K. and Uzuki, H. 1991. Physical distortion of a call of the jungle crow (*Corvus macrorhynchos*) in the process of propagation. – J. Ethol. 9: 147–148.

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs E. and Wagner, H. 2018. vegan: Community Ecology Package. R package version 2.5-3.

Olson, D. M., Dinerstein, E., Wikramanayake, E. D., Burgess, N. D., Powell, G. V. N., Underwood, E. C., D’amico, J. A., Itoua, I., Strand, H. E., Morrison, J. C., Loucks, C. J., Allnutt, T. F., Ricketts, T. H., Kura, Y., Lamoreux, J. F., Wettengel, W. W., Hedao, P. and Kassem, K. R. 2001. Terrestrial Ecoregions of the World: A New Map of Life on Earth. – Bioscience 51: 933.

Podos, J. and Warren, P. S. 2007. The Evolution of Geographic Variation in Birdsong. – Adv. Study Behav. 37: 403–458.

Ranjard, L., Anderson, M. G., Rayner, M. J., Payne, R. B., McLean, I., Briskie, J. V., Ross, H. A., Brunton, D. H., Woolley, S. M. N. and Hauber, M. E. 2010. Bioacoustic distances between the begging calls of brood parasites and their host species: A comparison of metrics and techniques. – Behav. Ecol. Sociobiol. 64: 1915–1926.

Ranjard, L. and Ross, H. A. 2008. Unsupervised bird song syllable classification using evolving neural networks. – J. Acoust. Soc. Am. 123: 4358–4368.

Rosa, G. L. M., Ellis, J. M., Bonaccorso, E. and Dos Anjos, L. 2016. Friend or foe? Social system influences the allocation of signals across functional categories in the repertoires of the New World jays. – Behaviour 153: 467–524.

Ridgely, R. S. and Tudor, G. 2009. Field guide to the songbirds of South America: the passerines. – University of Texas Press, p. 517.

Sarda-Espinosa, A. 2018. dtwclust: Time Series Clustering Along with Optimizations for the Dynamic Time Warping Distance. R package version 5.5.1.

Slabbekoorn, H. and Smith, T. B. 2002. Bird song, ecology and speciation. – Philos. Trans. R. Soc. B Biol. Sci. 357: 493–503.

Sumudu, F. W., Peterson, A. T. and Li, S. H. 2017. Reconstructing the geographic origin of the New World jays. – Neotrop. Biodivers. 3: 80–92.

Thomas, P. 2013. *Araucaria angustifolia*. – The IUCN Red List of Threatened Species 2013.

Veloso, H. P., Rangel-Filho, A. L. R. and Lima, J. C. A. 1991. Classificação da vegetação brasileira, adaptada a um sistema universal. – IBGE.

Venkatraman, M. X., Deraad, D. A., Tsai, W. L. E., Zarza, E., Zellmer, A. J., Maley, J. M. and McCormack, J. E. 2019. Cloudy with a chance of speciation: integrative taxonomy reveals extraordinary divergence within a Mesoamerican cloud forest bird. – Biol. J. Linn. Soc. 126: 1–15.

Williams, D. A. 2004. Female control of reproductive skew in cooperatively breeding brown jays (*Cyanocorax morio*). – Behav. Ecol. Sociobiol. 55: 370–380.

Williams, D. a. and Hale, A. M. 2006. Helper Effects on Offspring Production in Cooperatively Breeding Brown Jays (*Cyanocorax Morio*). – Auk 123: 847.

Williams, D. A. and Rabenold, K. N. 2005. Male-biased dispersal, female philopatry, and routes to fitness in a social corvid. – J. Anim. Ecol. 74: 150–159.

Wright, T. F. and Dahlin, C. R. 2017. Vocal dialects in parrots: Patterns and processes of cultural evolution. – Emu 118: 50–66.

Yandell, D. D., Hochachka, W. M., Pruett-Jones, S., Webster, M. S. and Greig, E. I. 2018. Geographic patterns of song variation in four species of *Malurus* fairy-wrens. – J. Avian Biol. 49: 1–15.

Yorzinski, J. L. and Vehrencamp, S. L. 2009. The Effect of Predator Type and Danger Level on the Mob Calls of the American Crow. – Condor 111: 159–168.

Young, S. J., Evermann, G., Gales, M. J. F., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., and Woodland, P. C. 2006. The HTK Book, version 3.4. – Cambridge University Engineering Department.

Zamudio, K. R., Bell, R. C. and Mason, N. A. 2016. Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. – Proc. Natl. Acad. Sci. 113: 8041–8048.

FIGURES

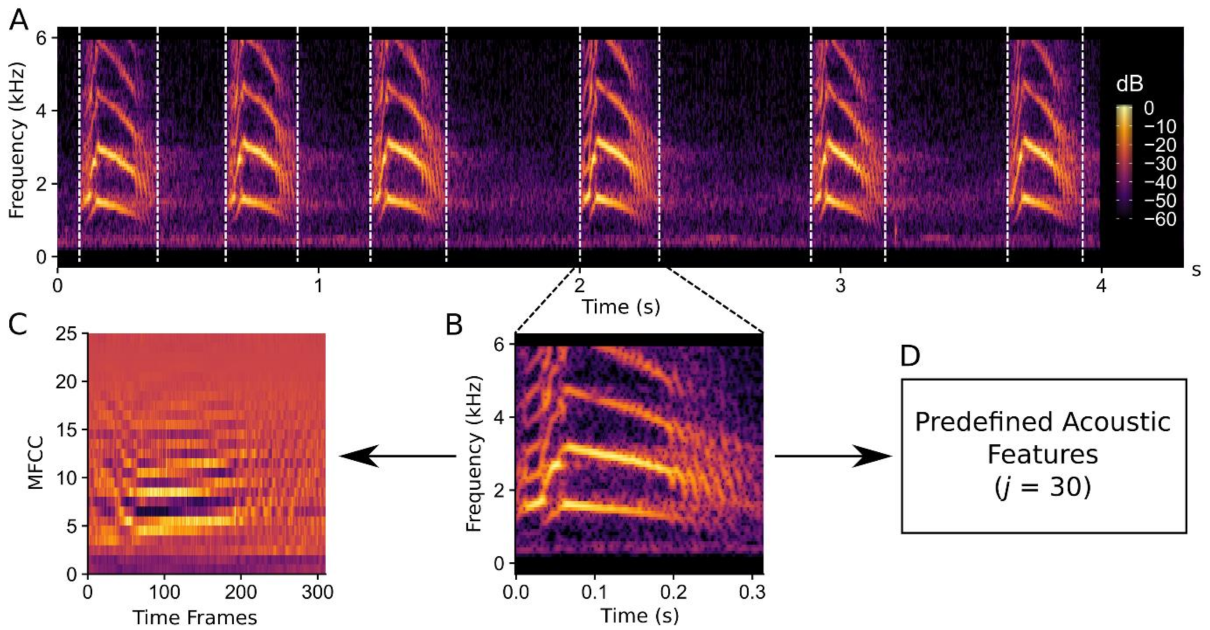


Figure 1. Sample selection and sound representations for acoustic analysis. (A) Sample selection of social-calls from recordings made in the field, start and end, indicated by dashed lines (window = Hann; frequency scale = Mel; fft size = 512; overlap = 90%). (B) A band-pass filter of 300-6000Hz followed by normalization to 0dB were applied to each resulting selection. (C) Sample representation by a set of 26 Mel-frequency Cepstral Coefficients (MFCC; win = 5ms, hop = 1ms, lift.exp = 22, pre.filt = 0.97), sampled from the 300-6000Hz frequency band. (D) Sample representation by Predefined Acoustic Features (PAF; 27 acoustic features as available in function ‘specan’ from the R package ‘warbleR’).

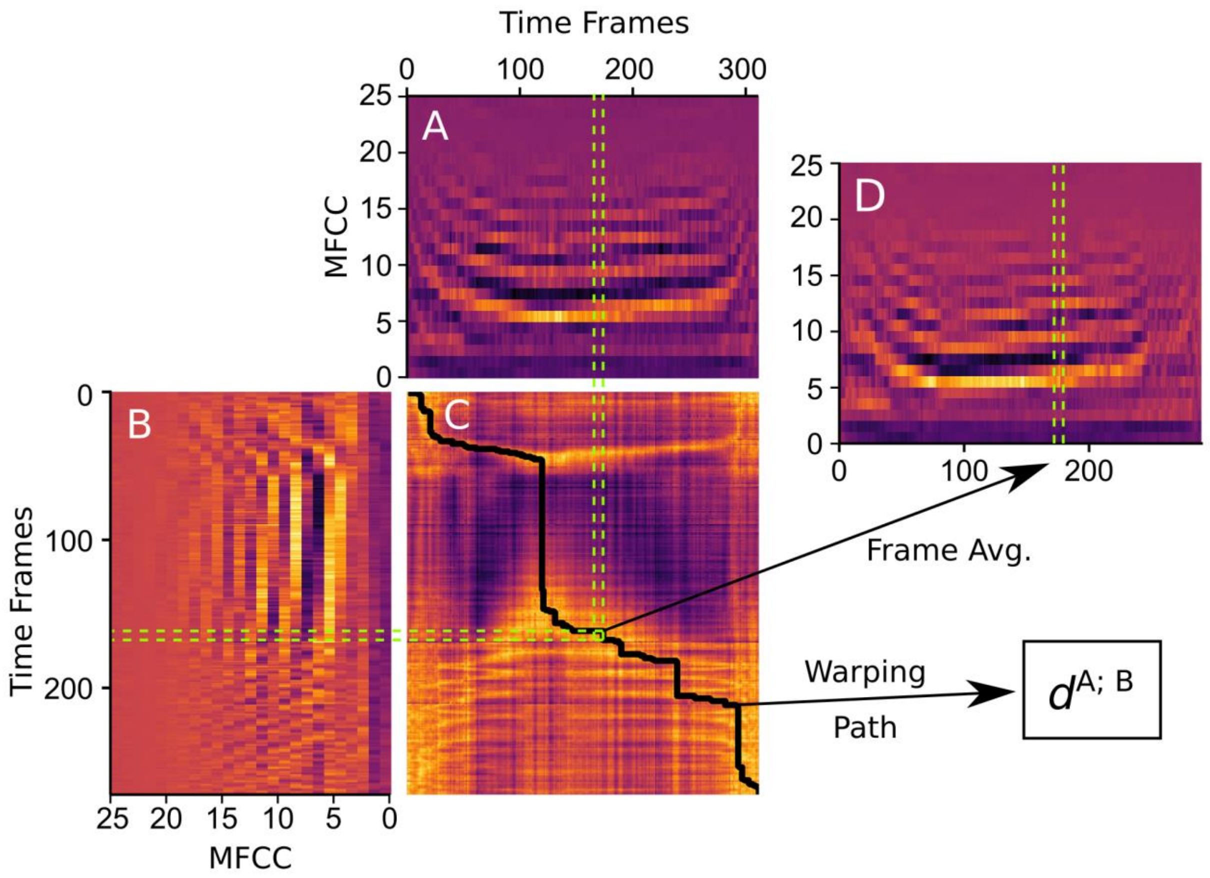


Figure 2. Demonstration of the Dynamic Time Warping (DTW) of two different sets of MFCC (A and B; lighter colour represents higher energy content). (C) Cost-matrix representing frames alignment from all frames from A and B, with lighter colours representing better alignments. The warping path (black solid line) indicates best match between frames and its total length represents the DTW distance, or vocal distance, between A and B. (D) An averaged sample can be produced by averaging the energy content of aligned frames (intervals between green dashed lines) in A and B, which is interpolated afterwards to their average number of frames.

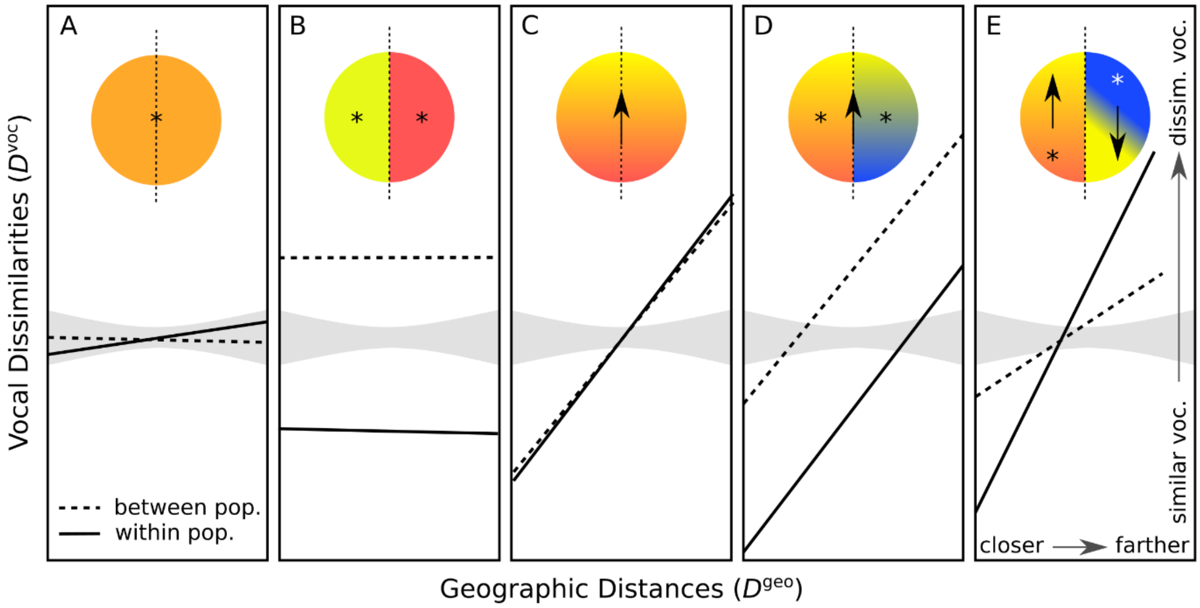


Figure 3. Partially competing hypothesis of geographic variation represented in linear effects of D^{geo} over D^{voc}, adapted from Yandell et al. (2018). To assess the effect of population sharing (D^{pop}), comparisons within populations are coded as a solid line and between populations as a dashed line. The hypotheses are: (A) Null effect, no geographic variation at all; (B) Population effect, though geographic distance has no effect, both populations diverge; (C) Geographic effect, with no divergence, both populations share the same geographic distance effects; (D) Non-interacting effect, though populations diverge, geographic distance effects are shared; (E) Interacting effect, populations diverge and have different geographic distance effects inside each population. Circles illustrate a spatial structure of phenotypic variation (colour coded) expected in each case.

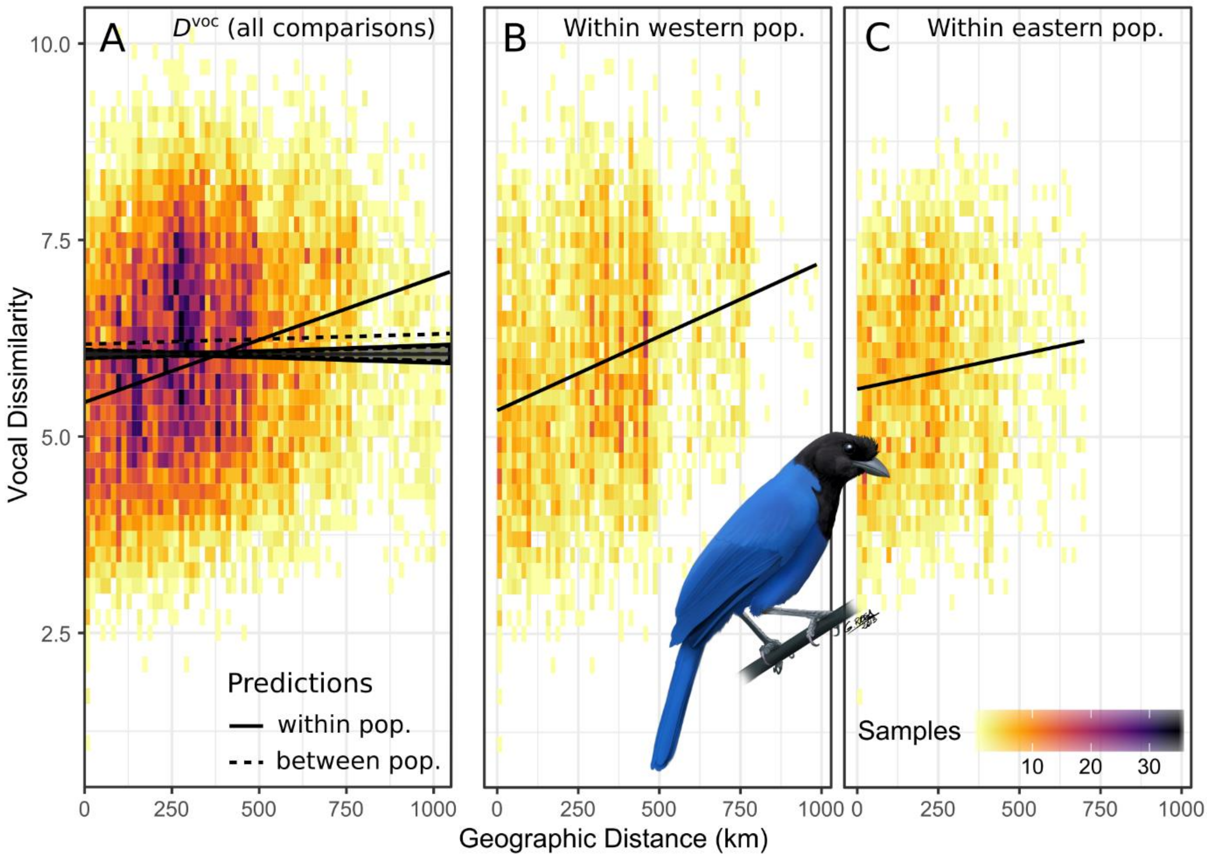


Figure 4. Linear effect of geographic distance (D^{geo}) over vocal distances (D^{voc}). Separate coefficients based on predictions from our data for comparisons within the same population are shown as continuous lines and between populations as dashed lines. Cell colours represent the density of pairwise comparisons. Shaded confidence intervals represent coefficients from empirical null distributions, calculated from the same model under 1000 randomizations of D^{voc}, also showing different comparison types. (A) A lower intercept and steeper slope of geographic distance on vocal dissimilarity was observed when comparisons were made within populations, than between populations, in a pattern consistent with the Interaction effect. (B and C) As confirmation of the interaction between geographic distance and population sharing.

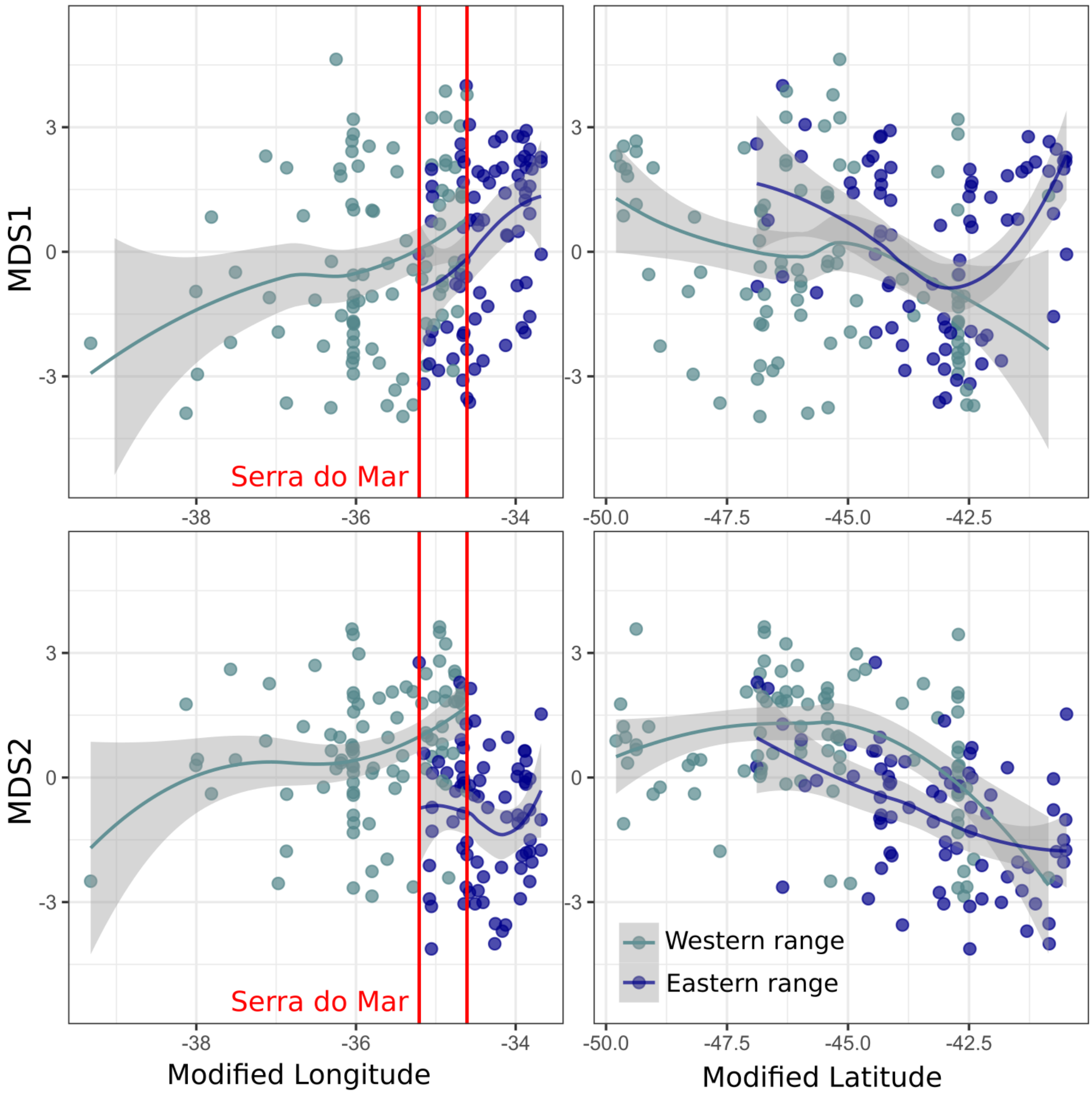


Figure 5. Directional trends of vocal distances represented by the first two NMDS axes of D_{voc} in geographic space. Geographic coordinates of the points were rotated by 0.4 radians, so the geographic separation between the two populations is represented orthogonally to the first axis, the modified longitude. The dashed lines represent a local polynomial regression (LOESS; span = 1) with 95% confidence intervals represented as shaded areas.

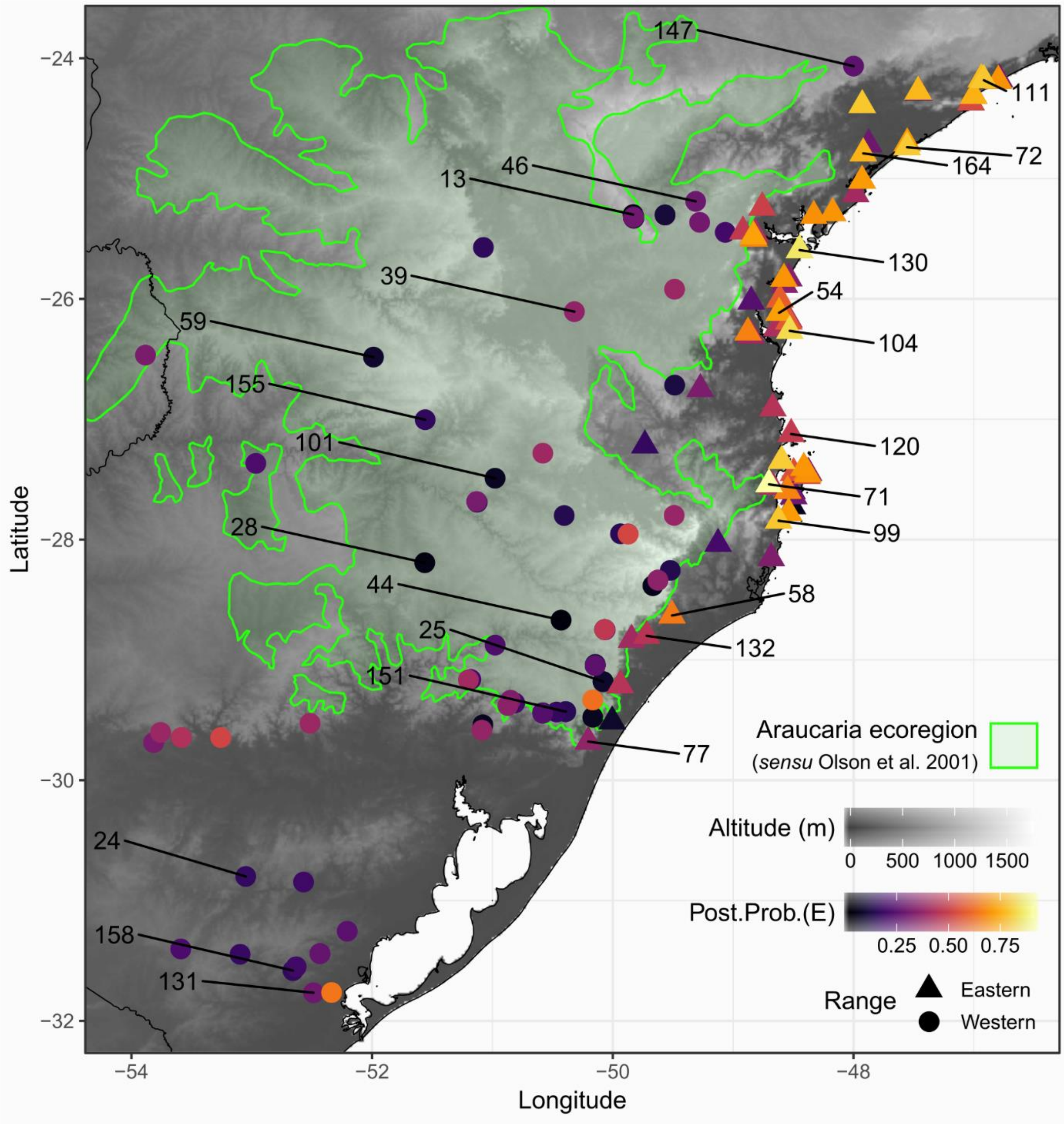


Figure 6. Geographic distribution of recordings indicating original population assignments (symbol shape) and average posterior probabilities of classification as eastern-like social-calls (symbol colour). Such predictions were made under a Random Forest classification model, in which new ranges were predicted from sample predefined acoustic features (PAF). Recording numbers are provided as reference for averaged spectrograms and recording metadata available in Figure S3 and Table S1. Altitude data was retrieved from WorldClim (Fick and Hijmans 2017) and limits of the Araucaria ecoregion from (Olson et al. 2001).

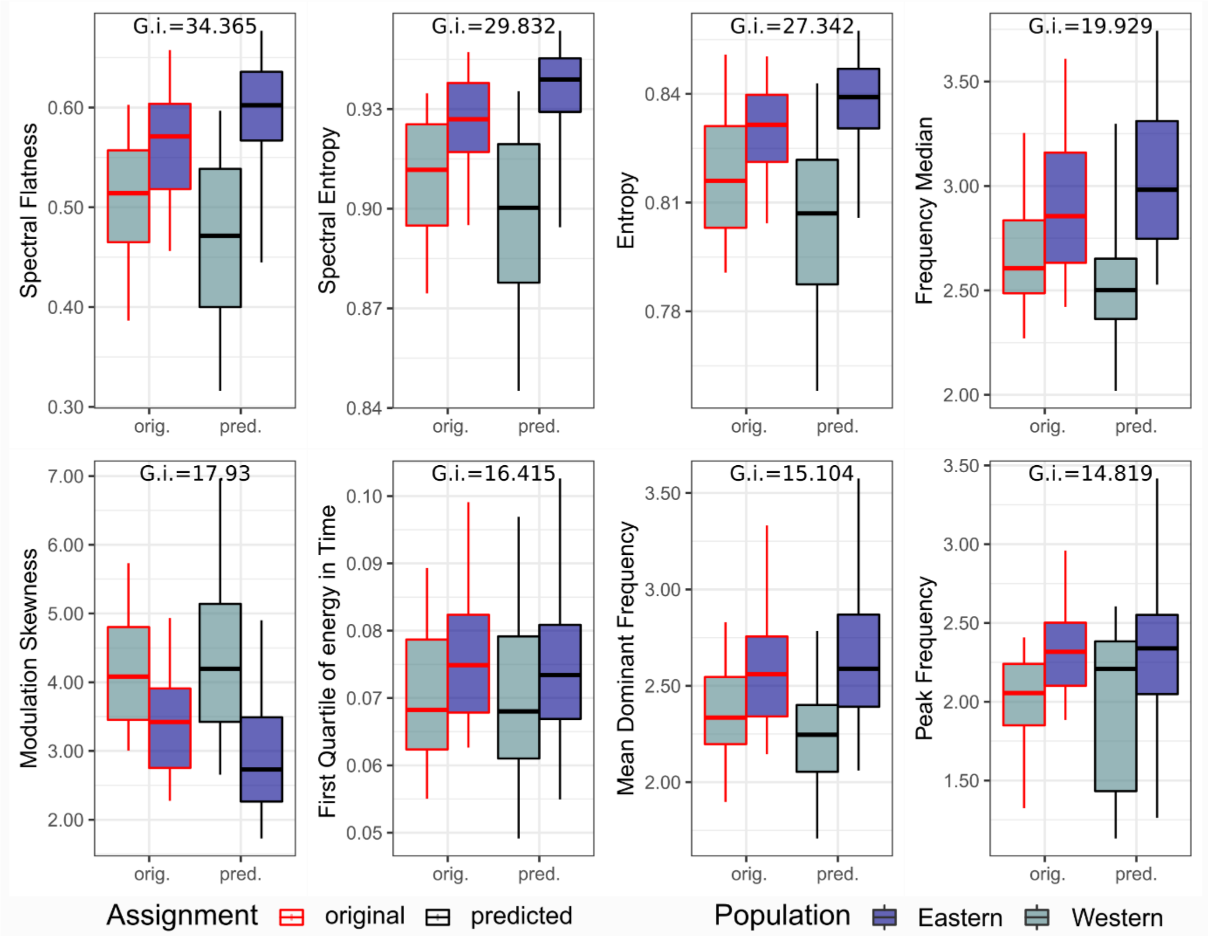


Figure 7. Comparison of the distributions of Predefined Acoustic Features (PAF) for eastern (cobalt blue) and western (greenish blue) populations under two alternative population assignments, the original (red boxes) and those based on predictions Random Forest model (black boxes). The eight acoustic features are those witch contributions for accurate classification (Gini Index). Vertical lines represent the 5th and 95th percentile intervals, boxes represent the 25th and 75th percentile intervals, and the horizontal line represent the mean.

TABLES

Table 1. Linear effects of geographic distance (*D^{geo}*), population sharing (*D^{pop}*), and their interaction (*D^{geo}* and *D^{pop}*) over vocal distances (*D^{voc}*). Regression coefficients were compared (two-tailed t-test) to the 95% confidence intervals of their empirical null distributions, generated in 1000 randomizations of *D^{voc}*. Coefficients out of their empirical null distributions are considered statistically different at the 5% level.

<table>
<tr><th colspan="5">Linear regression (R^2 = 0.04479; F3, 13691 = 214; DF13691 = 1.261)</th><th colspan="2">Comparison to Null distribution (DF = 999)</th></tr>
<tr><th>Coefficient</th><th>Estimate</th><th>Std. Error</th><th>t</th><th>Pr(>|t|)</th><th>t</th><th>P</th></tr>
<tr><td>Intercept</td><td>5.43584</td><td>0.02724</td><td>199.58284</td><td><0.00001</td><td>606.56890</td><td><0.00001</td></tr>
<tr><td>D^{geo}</td><td>0.00159</td><td>0.00008</td><td>19.43915</td><td><0.00001</td><td>-511.50620</td><td><0.00001</td></tr>
<tr><td>D^{pop}</td><td>0.73798</td><td>0.04141</td><td>17.82350</td><td><0.00001</td><td>-478.55140</td><td><0.00001</td></tr>
<tr><td>Interaction</td><td>-0.00146</td><td>0.00011</td><td>-13.48174</td><td><0.00001</td><td>376.67590</td><td><0.00001</td></tr>
</table>

SUPPLEMENTARY MATERIAL

Table S1. Metadata of the recordings analysed in this study, including recording source, catalogue number for external reference, number of retrieved samples, and geographic coordinates and original population assignments included in this study (E = Eastern; W = Western). Posterior probabilities for population assignment predictions (Pred.W and Pred.E) and the identification number for reference in this study (ID) are also provided. Predictions were not available for recording which entire content was used to train the classification model.

Source	Catalog. Nº	Samples	Long.	Lat.	Orig.Pop	Pred.W	Pred.E	ID
Anjos, L. dos	013-004	82	-49.827705	-25.322676	W	0.881	0.119	1
Anjos, L. dos	014-001	54	-49.827705	-25.322676	W	0.953	0.047	2
Anjos, L. dos	014-004	182	-49.565556	-25.301944	W	0.893	0.108	3
Anjos, L. dos	014-005	42	-49.827705	-25.322676	W	0.704	0.297	4
Anjos, L. dos	014-006	179	-49.827705	-25.322676	W	0.977	0.023	5
Anjos, L. dos	017-001	126	-49.827705	-25.322676	W	0.871	0.129	6
Anjos, L. dos	017-002	125	-48.816111	-25.460556	E	0.728	0.272	7
Anjos, L. dos	017-003	66	-48.816111	-25.460556	E	0.567	0.433	8
Anjos, L. dos	020-002	115	-49.827705	-25.322676	W	0.887	0.113	9
Anjos, L. dos	023-004	79	-49.827705	-25.322676	W	0.929	0.072	10
Anjos, L. dos	023-005	301	-49.827705	-25.322676	W	0.958	0.042	11
Anjos, L. dos	061-004	93	-49.825636	-25.327783	W	0.972	0.028	12
Anjos, L. dos	070-014-019	56	-49.831472	-25.299056	W	0.981	0.019	13
Anjos, L. dos	070-017-019	67	-49.827705	-25.322676	W	0.944	0.056	14
Anjos, L. dos	077-010	17	-49.826625	-25.3144	W	0.824	0.176	15
Anjos, L. dos	077-014	38	-49.826625	-25.3144	W	0.978	0.023	16
Anjos, L. dos	078-019-020	137	-49.565556	-25.301944	W	0.976	0.025	17
Anjos, L. dos	079-005-006	58	-49.826625	-25.3144	W	0.923	0.077	18
Anjos, L. dos	080-019	64	-49.825636	-25.327783	W	0.975	0.025	19
Anjos, L. dos	081-004	14	-49.825636	-25.327783	W	0.923	0.077	20
Anjos, L. dos	97	117	-48.3875	-27.470833	E	0.477	0.524	21
Anjos, L. dos	98	374	-50.403611	-27.8	W	0.861	0.139	22
Macaulay Library	19583	7	-52.9667	-27.3667	W	0.776	0.224	23
Macaulay Library	20076	15	-53.05	-30.8	W	0.832	0.168	24
Macaulay Library	34022	81	-50.080587	-29.180254	W	0.951	0.049	25
Macaulay Library	34025	56	-50.080587	-29.180254	W	0.912	0.088	26
Macaulay Library	57901	26	-47.98198	-25.129631	E	0.591	0.409	27
Macaulay Library	63638	389	-51.562477	-28.192477	W	0.951	0.049	28
Macaulay Library	90137	108	-49.6667	-28.3833	W	0.938	0.062	29
Macaulay Library	91074	93	-52.5697	-30.8464	W	0.809	0.192	30
Macaulay Library	135832	55	-53.8833	-26.4667	W	0.679	0.321	31
WikiAves	10006	3	-47.8739	-24.7144	E	0.704	0.296	32
WikiAves	44770	82	-53.0972	-31.4472	W	0.822	0.178	33
WikiAves	58873	18	-52.3369	-31.7647	W	0.358	0.643	34

WikiAves	63691	43	-47.4622	-24.2764	E	0.232	0.768	35
WikiAves	108418	29	-50.0675	-28.7475	W	0.893	0.107	36
WikiAves	181862	7	-51.1789	-29.1628	W	0.789	0.211	37
WikiAves	186635	13	-48.5475	-27.5944	E	0.572	0.428	38
WikiAves	240688	40	-50.3194	-26.1056	W	0.634	0.367	39
WikiAves	280472	46	-51.1269	-27.6897	W	0.875	0.125	40
WikiAves	333533	34	-49.8728	-27.9556	W	0.476	0.524	41
WikiAves	348342	3	-50.1464	-29.0472	W	0.744	0.257	42
WikiAves	357289	13	-46.7878	-24.1736	E	0.695	0.305	43
WikiAves	374884	22	-50.4294	-28.6694	W	0.960	0.041	44
WikiAves	412861	14	-48.3294	-25.3092	E	0.307	0.693	45
WikiAves	504381	6	-49.3114	-25.1892	W	0.712	0.288	46
WikiAves	507151	25	-48.5475	-27.5944	E	0.332	0.668	47
WikiAves	518855	3	-49.7331	-27.2172	E	0.827	0.173	48
WikiAves	578255	10	-48.6181	-26.1156	E	0.576	0.424	49
WikiAves	580030	25	-50.5828	-29.4403	W	0.887	0.113	50
WikiAves	598858	7	-50.5828	-29.4403	W	0.867	0.133	51
WikiAves	614925	94	-50.1464	-29.0472	W	0.813	0.187	52
WikiAves	615290	13	-50.1464	-29.0472	W	0.937	0.064	53
WikiAves	723847	6	-48.6181	-26.1156	E	0.277	0.723	54
WikiAves	724116	2	-50.5814	-27.2822	W	0.601	0.400	55
WikiAves	724181	48	-48.448327	-25.594151	E	0.256	0.744	56
WikiAves	731570	26	-50.0675	-28.7475	W	0.909	0.091	57
WikiAves	751327	4	-49.507758	-28.631108	E	0.352	0.649	58
WikiAves	771125	11	-51.9886	-26.4836	W	0.912	0.089	59
WikiAves	784716	12	-48.6703	-26.91	E	0.549	0.452	60
WikiAves	814270	25	-47.0011	-24.3119	E	0.245	0.755	61
WikiAves	860082	4	-48.575	-25.8817	E	0.654	0.346	62
WikiAves	875717	4	-50.0675	-28.7475	W	0.527	0.473	63
WikiAves	925611	17	-47.0011	-24.3119	E	0.439	0.561	64
WikiAves	938635	3	-53.8147	-29.6867	W	0.682	0.318	65
WikiAves	978827	3	-53.589667	-31.402632	W	0.784	0.217	66
WikiAves	989845	13	-50.814867	-29.357972	W	0.803	0.197	67
WikiAves	993848	6	-48.507771	-27.723063	E	0.941	0.059	68
WikiAves	1006482	24	-51.131194	-27.683777	W	0.684	0.316	69
WikiAves	1052695	3	-52.515249	-29.526553	W	0.606	0.394	70
WikiAves	1070329	41	-48.702049	-27.541191	E	0.089	0.911	71
WikiAves	1078098	3	-47.554317	-24.73924	E	0.199	0.801	72
WikiAves	1159883	7	-48.57753	-25.832072	E	0.298	0.703	73
WikiAves	1179327	8	-48.849409	-26.304518	E	0.600	0.400	74
WikiAves	1228005	19	-49.624978	-28.337947	W	0.871	0.129	75
WikiAves	1259090	14	-49.873211	-27.952686	W	0.706	0.294	76
WikiAves	1298496	12	-50.203395	-29.679137	E	0.632	0.368	77
WikiAves	1332940	20	-51.087619	-29.584159	W	0.624	0.376	78
WikiAves	1345071	13	-48.5375	-25.829167	E	0.655	0.345	79
WikiAves	1349463	5	-47.457104	-24.283929	E	0.717	0.283	80

WikiAves	1353203	12	-48.876555	-26.279126	E	0.333	0.667	81
WikiAves	1398138	4	-51.074945	-25.572163	W	0.841	0.158	82
WikiAves	1426618	5	-53.258692	-29.645663	W	0.480	0.520	83
WikiAves	1492064	35	-48.849842	-26.022076	E	0.756	0.245	84
WikiAves	1535335	32	-49.624978	-28.337947	W	0.898	0.102	85
WikiAves	1569643	11	-49.844118	-28.832517	E	0.593	0.407	86
WikiAves	1620470	4	-50.063613	-28.749906	W	0.798	0.203	87
WikiAves	1640114	27	-49.624978	-28.337947	W	0.728	0.272	88
WikiAves	1678897	2	-48.538349	-27.779353	E	0.285	0.715	89
WikiAves	1698331	28	-49.624978	-28.337947	W	0.638	0.362	90
WikiAves	1772445	15	-47.024574	-24.365941	E	0.714	0.286	91
WikiAves	1835181	3	-49.936514	-29.210649	E	0.660	0.340	92
WikiAves	1843570	8	-47.553137	-24.706954	E	0.286	0.714	93
WikiAves	1876720	15	-48.682866	-28.155477	E	0.666	0.334	94
WikiAves	1884167	13	-48.386199	-27.443601	E	0.343	0.657	95
WikiAves	1888139	40	-48.6125	-26.004167	E	0.410	0.590	96
WikiAves	1895658	4	-48.5375	-26.1875	E	0.421	0.580	97
WikiAves	1914446	8	-49.125091	-28.035606	E	0.793	0.207	98
WikiAves	1923168	23	-48.622334	-27.843421	E	0.204	0.796	99
WikiAves	1925075	6	-49.486379	-25.919466	W	0.611	0.389	100
WikiAves	1957348	89	-50.978151	-27.48997	W	0.927	0.073	101
WikiAves	1971027	5	-48.547637	-27.587796	E	0.558	0.442	102
WikiAves	1993552	14	-52.206794	-31.255277	W	0.748	0.252	103
WikiAves	2016892	20	-48.529208	-26.26569	E	0.167	0.833	104
WikiAves	2042731	2	-48.67415	-27.560023	E	0.565	0.436	105
WikiAves	2053787	10	-46.795833	-24.1875	E	0.308	0.692	106
WikiAves	2074301	7	-48.497289	-27.446715	E	0.450	0.550	107
WikiAves	2156520	29	-48.4125	-27.404167	E	0.295	0.705	108
WikiAves	2178199	33	-49.522575	-28.25648	W	0.860	0.140	109
WikiAves	2187389	9	-47.023852	-24.365999	E	0.466	0.534	110
WikiAves	2204056	6	-46.916862	-24.182287	E	0.202	0.798	111
WikiAves	2258880	6	-48.1725	-25.286944	E	0.298	0.702	112
WikiAves	2315076	2	-50.582108	-29.444098	W	0.761	0.240	113
WikiAves	2329626	4	-48.597522	-27.34659	E	0.194	0.806	114
WikiAves	2342517	6	-49.489834	-27.798671	W	0.616	0.384	115
WikiAves	2354233	11	-47.927217	-24.388604	E	0.206	0.794	116
WikiAves	2376794	5	-48.5375	-26.170833	E	0.586	0.414	117
WikiAves	2417230	11	-48.831539	-25.478482	E	0.562	0.438	118
WikiAves	2430855	14	-48.831539	-25.478482	E	0.271	0.730	119
WikiAves	2509363	24	-48.517739	-27.122836	E	0.515	0.485	120
WikiAves	2511742	45	-50.391498	-29.42988	W	0.884	0.116	121
WikiAves	2511903	3	-48.918734	-25.439351	E	0.523	0.478	122
WikiAves	2518564	4	-50.85012	-29.335507	W	0.670	0.331	123
WikiAves	2526816	23	-51.080088	-29.539342	W	0.931	0.069	124
WikiAves	2543610	3	-48.374219	-27.444116	E	0.666	0.334	125
WikiAves	2556342	17	-46.795833	-24.1875	E	0.299	0.702	126

WikiAves	2584969	2	-49.278276	-25.363293	W	0.693	0.307	127
WikiAves	2585301	9	-50.004781	-29.510438	E	0.913	0.087	128
WikiAves	2612429	10	-48.757968	-25.240193	E	0.600	0.400	129
WikiAves	2629686	22	-48.448327	-25.594151	E	0.143	0.857	130
WikiAves	2636267	16	-52.487008	-31.766839	W	0.707	0.294	131
WikiAves	2680911	38	-49.7188	-28.79972	E	0.539	0.461	132
WikiAves	2698278	25	-52.433302	-31.440644	W	0.724	0.276	133
WikiAves	2698279	35	-52.433302	-31.440644	W	0.831	0.169	134
WikiAves	2723654	1	-52.433302	-31.440644	W	***NA***	***NA***	***NA***
WikiAves	2740015	14	-47.928483	-25.016908	E	0.525	0.475	135
WikiAves	2750931	16	-53.585237	-29.64525	W	0.559	0.441	136
WikiAves	2774939	15	-48.757968	-25.240193	E	0.495	0.505	137
WikiAves	2784217	29	-52.630992	-31.550599	W	0.805	0.195	138
WikiAves	2786672	10	-47.928483	-25.016908	E	0.283	0.717	139
WikiAves	2816271	30	-50.166578	-29.476675	W	0.930	0.071	140
WikiAves	2833685	5	-48.637571	-26.244018	E	0.565	0.435	141
WikiAves	2851315	3	-48.826225	-25.499526	E	0.328	0.672	142
WikiAves	2856391	29	-50.977634	-28.877104	W	0.772	0.228	143
WikiAves	2865113	2	-49.272379	-26.74913	E	0.677	0.324	144
WikiAves	2865983	43	-50.871946	-29.379738	W	0.644	0.356	145
WikiAves	2876842	1	-50.213056	-29.662082	E	***NA***	***NA***	***NA***
WikiAves	2881888	37	-49.936514	-29.210649	E	0.545	0.455	146
WikiAves	2886227	10	-47.999396	-24.064961	W	0.741	0.259	147
WikiAves	2909095	13	-50.391498	-29.42988	W	0.850	0.150	148
WikiAves	2948090	11	-49.935833	-27.951944	W	0.850	0.150	149
WikiAves	2955827	24	-50.145421	-29.03803	W	0.920	0.080	150
WikiAves	2979943	1	-50.391498	-29.42988	W	0.933	0.067	151
WikiAves	2981738	5	-48.510261	-27.597532	E	0.765	0.236	152
WikiAves	2993829	14	-53.757849	-29.602888	W	0.609	0.391	153
WikiAves	3001937	13	-46.937166	-24.173106	E	0.225	0.775	154
WikiAves	3044856	9	-51.559097	-27.004092	W	0.836	0.164	155
WikiAves	3150095	16	-48.378468	-27.458378	E	0.303	0.697	156
WikiAves	3171705	1	-49.065306	-25.44655	W	0.798	0.202	157
WikiAves	3206499	2	-52.65778	-31.582203	W	0.836	0.164	158
Xeno-canto	1303	5	-51.2001	-29.1667	W	0.594	0.406	159
Xeno-canto	222029	4	-50.166667	-29.333333	W	0.374	0.626	160
Xeno-canto	252810	8	-50.4667	-29.4334	W	0.828	0.172	161
Xeno-canto	294707	36	-49.4856	-26.7166	W	0.879	0.122	162
Xeno-canto	298132	6	-48.497	-27.6393	E	0.701	0.299	163
Xeno-canto	433994	12	-47.9185	-24.7899	E	0.240	0.760	164

Table S2. Multivariate set of Predefined Acoustic Variables (PAF), selected from the implementation of the ‘specan’ function in the package ‘warbleR’ v.1.1.12 (Araya-Salas and Smith-Vidaurre, 2017).

Variable	Unit	Definition
duration	s	Length of the signal
meanfreq	Hz	Weighted average of frequency by amplitude
sd	Hz	standard deviation of frequency weighted by amplitude
freq.median	Hz	The frequency at which the signal is divided in two frequency intervals of equal energy
freq.Q25	Hz	The frequency at which the signal is divided in two frequency intervals of 25% and 75% energy respectively
freq.Q75	Hz	The frequency at which the signal is divided in two frequency intervals of 75% and 25% energy respectively
freq.IQR	Hz	Frequency range between ‘freq.Q25’ and ‘freq.Q75’
time.median	s	The time at which the signal is divided in two time intervals of equal energy
time.Q25	s	The time at which the signal is divided in two time intervals of 25% and 75% energy respectively
time.Q75	s	The time at which the signal is divided in two time intervals of 75% and 25% energy respectively
time.IQR	s	Time range between ‘time.Q25’ and ‘time.Q75’
skew	-	Asymmetry of the spectrum
kurt	-	Peakedness of the spectrum
sp.ent	-	Energy distribution of the frequency spectrum (pure tone ~ 0; noisy ~ 1)
time.ent	-	Energy distribution on the time envelope (pure tone ~ 0; noisy ~ 1)
entropy	-	Spectral entropy
sfm	-	Spectral flatness (pure tone ~ 0; noisy ~ 1)
meanfun	Hz	Average of fundamental frequency measured across the acoustic signal
minfun	Hz	Minimum fundamental frequency measured across the acoustic signal
maxfun	Hz	Maximum fundamental frequency measured across the acoustic signal
meandom	Hz	Average of dominant frequency measured across the acoustic signal
mindom	Hz	Minimum of dominant frequency measured across the acoustic signal
maxdom	Hz	Maximum of dominant frequency measured across the acoustic signal
dfrange	Hz	Range of dominant frequency measured across the acoustic signal
modindx	-	Modulation index, calculated as the cumulative absolute difference between adjacent measurements of dominant frequencies divided by the dominant frequency range (1 means the signals is not modulated)
startdom	Hz	Dominant frequency measurement at the start of the signal
enddom	Hz	Dominant frequency measurement at the end of the signal
dfslope	kHz/s	Slope of the change in dominant frequency through time
meanpeakf	Hz	Mean of the change in peak frequency through time
peakf	Hz	Frequency with the highest energy

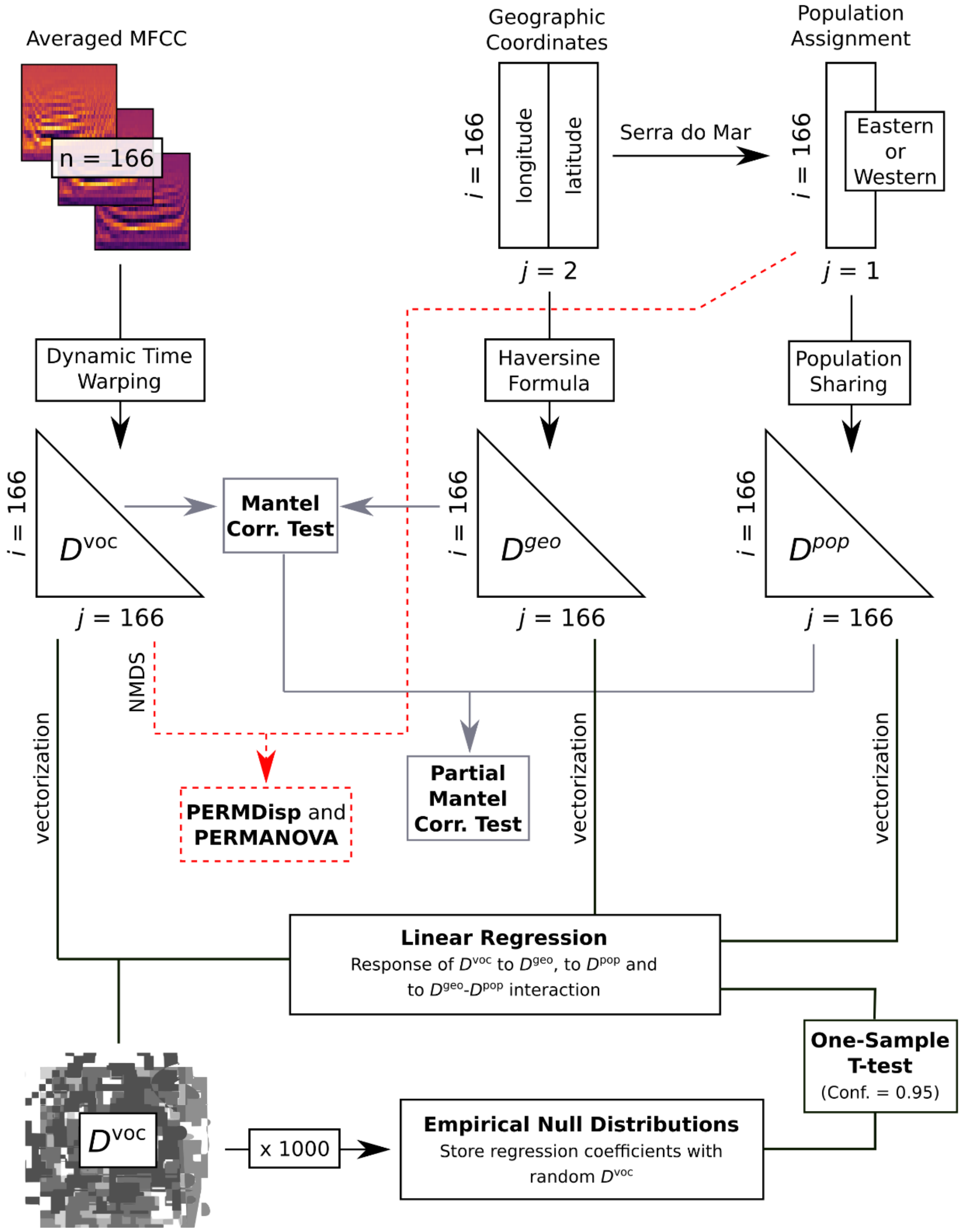


Figure S1. Schematic representation of statistical procedures, including data pre-processing to represent D^{voc}, D^{geo} and D^{pop}. Represented statistical procedures are the following: dispersion and location effects of population assignments over D^{voc} (red dashed lines); Mantel Correlation test and its partial alternative between pairwise comparisons (solid grey lines); and the regression to test the linear of D^{geo}, D^{pop} and the interaction between D^{geo} and D^{pop} with D^{voc}, including comparison to the 95% confidence interval of empirical null distributions taken from each coefficient after 1000 randomizations of D^{voc}.

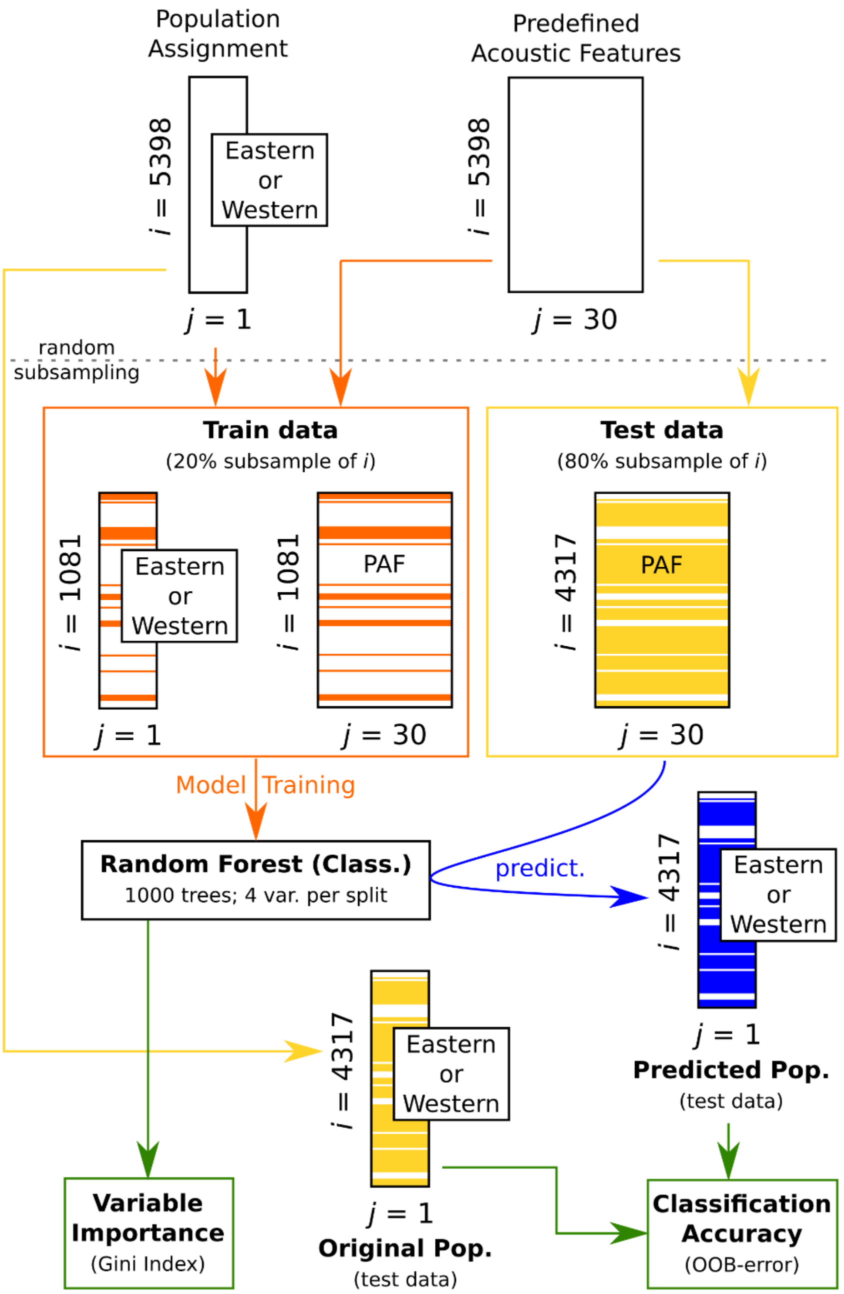


Figure S2. Schematic representation of Random Forest Classification. The model was initially trained with 20% of the samples and their population assignments (train data; orange). Using the test data (80%; yellow), including only PAF, new population assignments were predicted (blue). From the trained classification model, the importance of each PAF was compared with the Gini Index, in which higher values of a given acoustic feature represent larger contribution to the divergence between populations. The out of bag error rate (OOB) was estimated by comparing predicted population assignments with the original ones, and the higher the agreement between both, the lower the model error is.

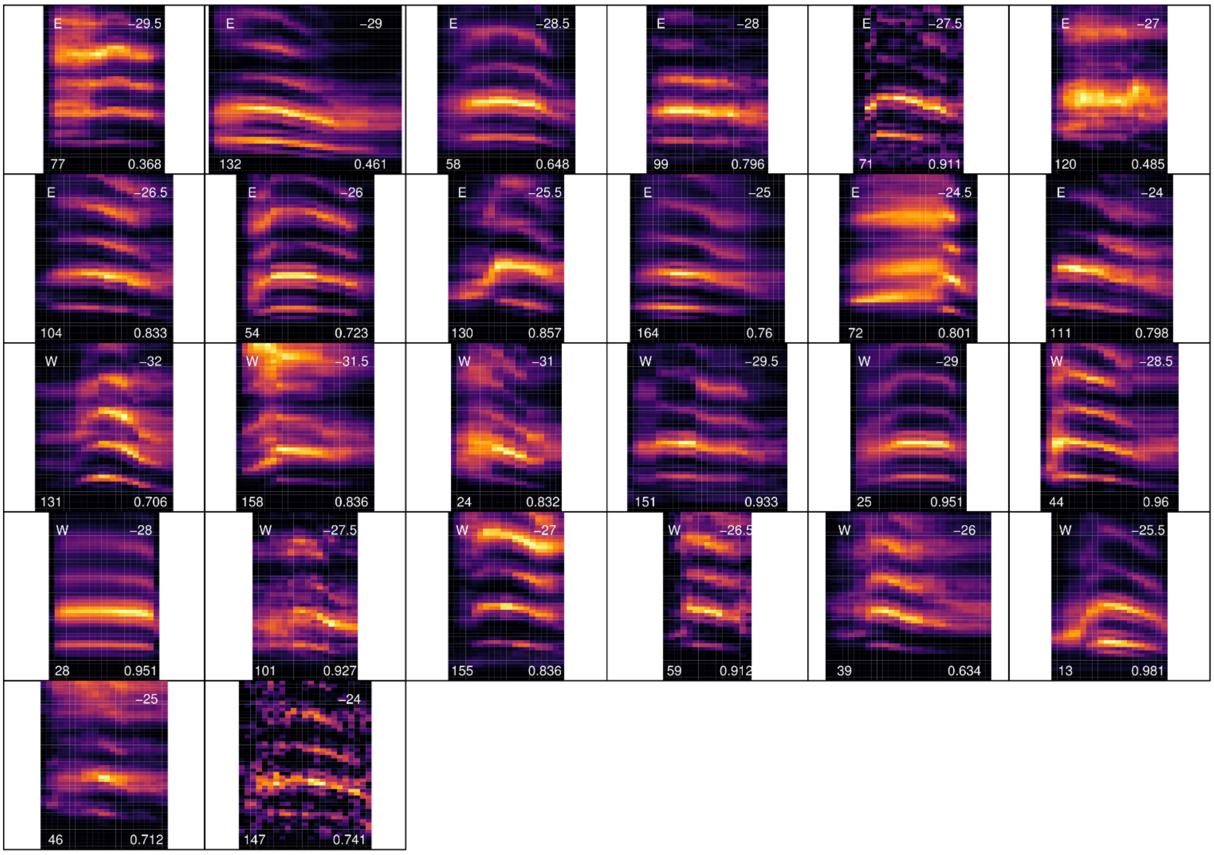


Figure S3. Averaged spectrograms of recordings (ID in bottom-left) with the highest posterior probabilities (bottom right) to be correctly assigned to their original populations (top-left), for each latitude window of half decimal degree in the species distribution (top-right). The averaging process is the same as described in Figure 2.

CONSIDERAÇÕES FINAIS

Por meio do capítulo primeiro desta tese, demonstrei como avaliar resultados de análises exploratórias produzidos a partir de diversas combinações de métodos. Apesar de não ser possível estabelecer um protocolo único para análises de agrupamento em sua miríade de aplicações, pude identificar condições que geram resultados indesejáveis para o estudo da estrutura de repertórios vocais. Além de oferecer sugestões para resultados mais objetivos e replicáveis, demonstro uma estratégia para representar características de repertórios vocais, de forma que comparações possam ser feitas mesmo entre resultados produzidos de forma independente. Não avaliei todas as possibilidades de forma exaustivas, logo, a ampliação dos diagnósticos realizados neste estudo com a inclusão de novos métodos pode aperfeiçoar recomendações para novos estudos.

No capítulo segundo, demonstrei como inferências estatísticas e predições por meio de aprendizado de máquina podem ser complementares para produzir conhecimento a partir de um padrão biológico complexo. Confirmei a hipótese de que vocalizações da gralha-azul, *Cyanocorax coeruleus*, variam no espaço e entre populações a Leste e Oeste da Serra do Mar. Mais estudos são necessários para verificar se esta divergência entre populações corresponde a diversidade genética antes desconhecida. Caso haja de fato uma correlação, será constatado não somente o valor das análises que produziram estes resultados, mas também reforçado o grande valor de coleções de sons para o estudo e conservação da Biodiversidade. Além disso, tal diversidade genética justificaria uma ampla reavaliação das atuais estratégias de conservação desta espécie. Ressaltando sua proteção prevista por lei no Estado do Paraná (Lei 7957 de 21 de novembro de 1984, Diário Oficial no. 1913 de 22 de novembro de 1984 no Diário Oficial do Estado do Paraná) que inclui grande parte de sua distribuição geográfica.

Oferecendo novas ferramentas, encorajo novos estudos sobre as gralhas do gênero *Cyanocorax* dado o potencial desta linhagem de corvídeos como modelos úteis para compreender padrões e processos biológicos ligados à evolução de comunicação vocal.