Universidade Federal de Uberlândia

Instituto de Biologia

Programa de Pós-graduação em Ecologia e Conservação dos Recursos Naturais

UMA NOVA ABORDAGEM PARA MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES A PARTIR DE REGISTROS RESTRITOS: APLICAÇÃO EM PROJEÇÕES DE PADRÕES FUTUROS DE FORMIGAS DO CERRADO

Rafaella Almeida Silvestrini

2019

Rafaella Almeida Silvestrini

UMA NOVA ABORDAGEM PARA MODELAGEM DE DISTRIBUIÇÃO DE ESPÉCIES A PARTIR DE REGISTROS RESTRITOS: APLICAÇÃO EM PROJEÇÕES DE PADRÕES FUTUROS DE FORMIGAS DO CERRADO

Tese apresentada à Universidade Federal de Uberlândia, como parte das exigências para obtenção do título de Doutor em Ecologia e Conservação de Recursos Naturais

Orientador:

Prof. Dr. Heraldo L. Vasconcelos

UBERLÂNDIA

Junho, 2019

ATA DE DEFESA

Programa de Pós-Graduação em:	Ecologia e Conservação de Recursos Naturais
Defesa de:	Tese, número 62, PPGECRN

Data:	vinte e oito de junho de dois mil e dezenove	Hora de início:	08:00	Hora de encerramento:	12:45

Matrícula do Discente:	11513ECR005
Nome do Discente:	Rafaella Almeida Silvestrini
Título do Trabalho:	Uma nova abordagem para modelagem de distribuição de espécies a parir de registros restritos: aplicação em projetos de padrões futuros de formigas de cerrado
Área de concentração:	Ecologia
Linha de pesquisa:	Ecologia de comunidades e ecossistemas
Projeto de Pesquisa de vinculação:	Padrões da biodiversidade e processos ecológicos em ecossistemas de Cerrado na região do Triângulo Mineiro e Sudeste de Goiás (sub-bacia do Rio Paranaíba).

Reuniu-se na Sala 102, Bloco 8C, Campus Umuarama, da Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ecologia e Conservação de Recursos Naturais em 03 de junho de 2019, assim composta: Doutores: Celine de Melo - UFU; Natália Mundim Tôrres - UFU; Daniel de Paiva Silva - IF Goiano; Paulo de Marco Junior - UFG e Heraldo Luis de Vasconcelos - UFU, orientador(a) do(a) candidato(a).

Iniciando os trabalhos o(a) presidente da mesa, Dr(a). Heraldo Luis de Vasconcelos, apresentou a Comissão Examinadora e o candidato(a), agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ulimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovado(a).

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação perinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Referência: Processo nº 23117.072130/2019-51

SEI nº 1466684

Dedicatória

Dedico este trabalho à Carla, que, por sua história de vida,
me mostrou que seria possível.

Agradecimentos

Agradeço ao Heraldo Vasconcelos, à UFU, aos meus professores, à CAPES, e aos colegas de laboratório, em especial à Renata Pacheco.

Agradeço também a todas as pessoas que colaboraram no meu processo, em especial aos meus pais, ao Henrique e à Flavia, que assumiram grande parte das minhas responsabilidades na minha ausência.

ÍNDICE

Resumo

Mudanças climáticas atreladas às modificações nos ambientes naturais têm impulsionado o uso de modelos de distribuição de espécies voltados ao planejamento da conservação. Contudo, o uso destes modelos tem sido limitado para espécies cujas informações geográficas são restritas, fato comum em espécies raras ou endêmicas, categorias especialmente relevantes sob o ponto de vista da preservação. Os insetos, um grupo funcional importante e que compõem a maior parcela da biodiversidade terrestre, raramente são considerados em estudos de conservação. Neste contexto, esta tese se propôs a: 1) avaliar técnicas de modelagem que sejam menos sensíveis à condição de restrição de dados, caracterizada tanto pela quantidade quanto pela abrangência geográfica amostral; 2) aplicar estas metodologias para predizer o padrão atual e futuro da distribuição de espécies de formigas do *Cerrado* e outras vegetações savânicas da América do Sul. O primeiro item foi alcançado comparando-se modelagens resultantes da utilização de dados restritos *versus* aquelas resultantes da totalidade dos dados disponíveis. Estas comparações foram realizadas para três espécies de formigas – com diferentes padrões geográficos e quantidades de registros de ocorrência – usando duas técnicas de modelagem – Maxent e Pesos de Evidência. Pesos de Evidência é um método Bayesiano já bastante utilizado na modelagem ambiental, mas que nunca foi aplicado à modelagem de distribuição de espécies com sucesso. Para adequar este método aos dados disponíveis, foi realizado um controle no número e na localização das pseudo-ausências durante a calibração. Condições de restrição de dados apresentaram entre 5 e 27 ocorrências restritas ao *Cerrado*, enquanto condições de dados amplos mostraram entre 16 e 47 registros distribuídos ao longo da América do Sul. As performances de cada experimento foram avaliadas por meio de: 1) inspeção visual dos mapas de probabilidade; 2) porcentagem de acerto de presenças e ausências, avaliadas por meio do TSS (do inglês, True Skill Statistics) calculado em diferentes escalas. O método Pesos de Evidência forneceu resultados significativamente melhores que o Maxent, principalmente quando o objetivo foi estimar a distribuição da espécie fora da área de abrangência da calibração. Então, esse método foi aplicado na segunda etapa da tese, para modelar a distribuição atual e futura de 12 espécies de formigas típicas das vegetações savânicas da América do Sul. Distribuições estimadas mostraram que, sob o pior cenário, estas espécies poderão

perder entre 37% e 88% de suas áreas de adequabilidade estimadas. O estudo sugeriu *Linepithema cerradense* como a espécie de menor grau de ameaça. Por outro lado, *Blepheridatta conops* e *Pheidole cyrtostela* podem ser as mais vulneráveis, com menos de 16% de suas áreas favoráveis atuais a serem preservadas em 2070. As demais espécies apresentaram taxas de preservação das áreas de adequabilidade atual entre 17% e 32%. Estima-se que as espécies possuam, em média, 13% da área favorável atual coincidindo com áreas de proteção, valor que pode chegar a 5% em 2070.

Abstract

The joint effect of changes in climate and on natural environments has triggered the use of species distribution models for conservation planning. However, application of these models has been very limited for species whose geographic information is restricted, a common fact in rare or endemic species, categories which are especially relevant from the point of view of preservation. Insects, an important functional group that compose the largest portion of terrestrial biodiversity, are rarely considered in conservation studies. In this context, the present thesis aims to: 1) evaluate modeling techniques that are less sensitive to the data restriction condition, characterized by both the quantity and the geographic range of the sample; 2) to apply these methodologies to predict the current and future pattern of the species of ants typical of the *Cerrado* and other types of savanna vegetation of South America. The first goal was achieved by comparing models resulting from the use of restricted data versus those resulting from the use of all data. These comparisons were made for three ant species - which have different geographic patterns and quantities of occurrence records - using two modeling techniques - Maxent and Weights of Evidence. Weights of Evidence is a Bayesian method widely used in environmental modeling, but that has never been successfully applied to modeling species distribution. In order to adapt this method for the application in our available data, I have controlled the number and location of pseudo-absences during the calibration phase. For each species, data restriction conditions comprised between 5 and 27 occurrences, restricted within the *Cerrado* Biome, while broad data varied between 16 and 47 records distributed throughout South America. The performances of each experiment were assessed by: 1) visual inspection of probability maps; 2) TSS (True skill statistics) evaluated according to different scales. The Weights of Evidence method provided significantly better results than Maxent, especially when the goal was to estimate the distribution of the species outside the range of the calibration area. Then, given the best performance of the Weights of Evidence method, it was applied to model current and future distribution of 12 species of ants typical of savanna vegetation in South America. Estimated distributions have shown that under the worst scenario, these species will have lost between 37% and 88% of their current estimated range. This study suggested that *Linepithema cerradense* is the least threatened species. On the other hand, *Blepheridatta*

conops and *Pheidole cyrtostela* might be the most vulnerable ones, with less than 16% of their current estimated distribution to be preserved in 2070. The other species may preserve, by 2070, between 17% and 32% of their current estimated distributions. These species might have, on average, 13% of their current suitable area overlapped with protection zones, a value that can decrease to only 5% by 2070.

Introdução Geral

O aquecimento global e as mudanças na cobertura vegetal constituem uma das maiores ameaças à conservação da biodiversidade. Estudos indicam que grande parte da América do Sul terá sua temperatura aumentada em mais de 4°C até o fim do século XXI (Magrin et al., 2014) e que o recente aquecimento já pode ter sido a causa da extinção de algumas espécies (Pounds et al., 2006; Urban, 2015). Por sua vez, as alterações antrópicas na paisagem natural podem impedir indivíduos de colonizarem regiões com climas propícios à manutenção de sua espécie (Marini et al., 2009). Frente a estas perspectivas, a biodiversidade do Cerrado, segundo maior bioma da América do Sul (MMA, 2019), pode estar seriamente ameaçada. O *Cerrado* possui apenas 2.85% do território sob proteção integral (MMA, 2019), aproximadamente metade de sua vegetação nativa já foi perdida e estima-se que apenas 19% da cobertura vegetal permanece inalterada (Strassburg et al., 2017). Estimativas sobre a diversidade do *Cerrado* apontam o bioma como a savana mais diversa do mundo, com alto grau de endemismos (Myers et al., 2000).

Neste contexto, a modelagem de distribuição de espécies surge como uma ferramenta essencial para planejar estratégias de conservação, uma vez que ela permite inferir sobre os locais propícios à manutenção de espécies frente às mudanças climáticas e as alterações na paisagem natural (Soberón e Peterson, 2004). Tal método tem como princípio básico predizer, por meio de fundamentos da ecologia e técnicas estatísticas e/ou de modelagem computacional, o padrão da distribuição potencial das espécies através de uma série de variáveis de entrada (Soberón e Nakamura, 2009), geralmente associadas ao clima e à estrutura da vegetação. O objetivo é estimar parcialmente o nicho da espécie (Peterson e Soberon, 2012) baseado em variáveis ambientais e desconsiderando outros fatores que também delineiam o nicho, como exemplo, relações entre as espécies dentro de sua comunidade e estrutura demográfica da população.

Apesar da vasta aplicação do método em espécies de interesse para conservação, como mamíferos e aves (e.g.: Marini et al., 2009; Ferraz et al., 2009; Zeilhofer et al., 2014; Charbonnel et al., 2016), o uso deste tipo de modelagem tem sido bastante limitado em espécies raras ou endêmicas - categorias especialmente importantes do ponto de vista da conservação - devido à falta de habilidade dos métodos em lidar com poucos registros de ocorrência (Lomba et al. 2010; Galante et al., 2017), uma condição frequente destas espécies.

Dada a inviabilidade de esperar o surgimento de informações adicionais destes grupos, torna-se urgente o desenvolvimento de técnicas capazes de superar o desafio da restrição de dados. Estudos recentes envolvendo esta temática mostraram que, dentre os métodos comumente utilizados em previsão de distribuição de espécies, o algoritmo Maxent é menos sensível à quantidade de registros (Wisz et al., 2008; Muscarella et al, 2014), se utilizado concomitantemente com técnicas de jacknife ou bootstrap, disponibilizadas no pacote ENMeval do R (Muscarella et al., 2014; Galante et al., 2017).

Recentemente, considerando novos métodos que podem ser eficazes na modelagem espacial com restrição de dados, as técnicas Bayesianas têm se destacado devido, principalmente, a suas poucas exigências em relação às hipóteses iniciais, o que as tornam facilmente adaptáveis a diversas situações. Dentre eles, duas técnicas são aplicadas em modelagem ambiental: as Redes Bayesianas e os Pesos de Evidência. As Redes Bayesianas têm aplicações bem recentes, destacando-se seu uso em modelagem de distribuição de espécies (Hamilton et al., 2015; Trifonova et al., 2017). O método baseia-se no uso de informações adicionais - tais como conhecimentos *à priori* das preferências da espécie - como forma de compensar a situação de poucos dados. Contudo, sua aplicação é bastante desafiadora pois, além de exigir conhecimentos detalhados a respeito da espécie, muitas vezes não disponíveis, ele requer um elevado grau de conhecimento estatístico e computacional do modelador por não se encontrar implementado em nenhum software de modelagem. Por outro lado, o método Pesos de Evidência (Bonham-Carter, 1994), apesar de nunca ter sido aplicado na modelagem de distribuição de espécies, é bem estabelecido no campo da modelagem ambiental, com aplicações em modelagens de desmatamento (Soares-Filho et al., 2006), incêndios florestais (Silvestrini et al., 2011) e ocorrências de minerais (Agterberg e Bonham-Carter, 1990). Semelhantemente aos outros métodos de previsão de distribuição de espécies, os Pesos de Evidência usam uma série de variáveis para criar relações e estimar pesos, que indicam como cada variável influencia na presença do fenômeno em estudo. Ao contrário das redes bayesianas, que apresentam dificuldade de implementação e exigem informações adicionais muitas vezes não disponíveis, as únicas exigências dos pesos de evidencia é que as variáveis explicativas apresentem correlação com o fenômeno em estudo e que sejam independentes e não correlacionadas entre si. Além destas vantagens, o fato deste método já se encontrar implementado em software de modelagem gratuito e de fácil acesso ao modelador (Dinamica EGO, Soares-Filho et al., 2013), o torna uma potencial metodologia para modelagem de espécies.

Independentemente da técnica utilizada, um ponto crucial em qualquer modelagem é o controle e a minimização dos erros (Beale e Lenon, 2012), que se inicia com a seleção de dados confiáveis de presença e/ou ausência da espécie, passa pela escolha de um algoritmo adequado para a quantidade de dados disponíveis e finaliza por meio de uma validação transparente. Tanto a escolha do algoritmo, a qual constitui a principal fonte de incerteza no processo de modelagem (Watling et al, 2015), quanto as técnicas de validação podem ser facilmente controladas pelo pesquisador. Por outro lado, a qualidade dos dados de ocorrência é o ponto mais difícil de ser atingido devido à possibilidade de coletas viciadas, as quais potencializam a chance de serem gerados padrões de distribuição espaciais errôneos, principalmente na situação de poucos dados (Boria et al., 2014). Uma modelagem confiável e eficaz é diretamente proporcional à qualidade dos dados de entrada.

A presente tese utiliza um banco de dados de ocorrência de espécies de formigas nas savanas da região do *Cerrado* (Vasconcelos et al. 2014, 2017), juntamente com outros dados disponíveis na base "Antmaps.org" (Janicki et al., 2016), para investigar métodos capazes de lidar com dados restritos e, então, modelar a distribuição espacial das espécies mais típicas do *Cerrado*. Para tanto, este trabalho foi dividido em dois capítulos.

O primeiro deles aborda o aspecto mais teórico da tese e está em processo de revisão na revista PeerJ. Nele, os métodos Maxent e Pesos de Evidência são comparados em relação à sua capacidade de lidar com a restrição de dados, seja ela pela pequena quantidade de registros ou pela restrição espacial da área de abrangência em que estes registros se encontram.

A segunda parte aplica o método Pesos de Evidência, que se revelou mais adequado para a situação de restrição de dados, na modelagem de 12 espécies de formigas que ocorrem predominantemente ou exclusivamente no *Cerrado* ou em outras áreas de vegetações savânicas da América do Sul. Com os modelos de distribuição das áreas de adequabilidade para cada espécie, determinei: (a) qual a porcentagem da área com adequabilidade ambiental atual de cada espécie que se encontra sob proteção, e (b) qual o tamanho das áreas favoráveis para cada espécie na atualidade e em 2070. Esta análise está sendo preparada para submissão na revista Journal of Insect Conservation.

Assim, pretende-se que este estudo, além de contribuir para o avanço de técnicas de modelagem de espécies raras ou endêmicas em geral, colabore com a inclusão dos insetos em estudos de planejamento da conservação, pois, apesar de representar a maior parcela da biodiversidade e constituir um grupo funcional importante (Edwards et al.,

1993; Thomas et al., 2008), eles raramente são levados em consideração devido à pequena disponibilidade de dados (Jenkins et al., 2013).

Referências

Beale CM, Lennon JK (2012) Incorporating uncertainty in predictive species distribution modelling. Philosophical Transactions of the Royal Society B, 367, 247-258.

Boria RA, Olson LE, Goodman SM, Anderson RP (2014) Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. Ecological Modelling, 275, 73-77.

Charbonnel A, Laffaille P, Biffi M, Blanc F, Maire A, Némoz M, et al. (2016) Can Recent Global Changes Explain the Dramatic Range Contraction of an Endangered Semi-Aquatic Mammal Species in the French Pyrenees? PLoS ONE 11(7): e0159941.

Edwards PJ, May RM, Webb NR (1993) Large-scale ecology and conservation biology, 1st edn. Blackwell Science Ltd., Oxford.

Ferraz KMPAMB, Siqueira MF, ALexandrino ER et al (2012) Environmental suitability of a highly fragmented and heterogeneous landscape for forest bird species in south-eastern Brazil. Environmental Conservation, 39 (4): 316–324.

Ferreira ME, Ferreira LG, Miziara F, Soares-Filho BS (2013) Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation. Journal of Land Use Science, 8:Issue 4.

Janicki J, Narula N, Ziegler M, Guénard B, Economo EP (2016) Visualizing and interacting with large-volume biodiversity data using client-server web-mapping applications: The design and implementation of antmaps.org. Ecological Informatics 32:185-193.

Lomba A, Pelissier L, Randim C et al (2010) Overcoming the rare species modelling paradox: A novel hierarchical framework applied to an Iberian endemic plant. Biological Conservartion 143: 2647:2657.

Magrin GO, Marengo JA, Boulanger JP et al (2014) Central and South America. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R. Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1499-1566.

Marini MA, Barbet-Massin M, Lopes LE, Jiguet F (2009) Major current and future gaps of Brazilian reserves to protect Neotropical savanna birds. Biological Conservation, 142:3039–3050.

MMA (Ministério do Meio Ambiente) (2019) http://www.mma.gov.br/biomas/*Cerrado*. Acesso em 27/05/2019

Myers N, Mittermeier R A, Mittermeier CG, Da Fonseca GA, Kent J (2000) Biodiversity hotspots for conservation priorities. Nature, 403(6772): 853-858.

Soares-Filho BS, Rodrigues HO, Costa WL (2013). Modeling Environmental Dynamics with Dinamica EGO. Available from: http://www.csr.ufmg.br/dinamica. (Acessed May, 2017).

Soberon J, Nakamura M (2009) Niches and distributional areas: Concepts, methods, and assumptions. PNAS, 106: 19644–19650.

Soberón J, Peterson AT (2004) Biodiversity informatics: Managing and applying primary biodiversity data. Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 359:689-698.

Soberón J, Peterson AT (2012) Species distribution modelling and ecological niche modelling: getting the concepts right. Natureza e Conservação 10(2):1-6

Zeilhofer P, Cezar A, Tôrres, NM, de Almeida Jácomo AT, Silveira L (2014) Jaguar Panthera onca Habitat Modeling in Landscapes Facing High Land-use Transformation Pressure - Findings from Mato Grosso, Brazil. Biotropica, 46: 98–105.

A comparison between Maxent and Weights of Evidence in modeling species with few and geographically restricted data

Abstract

Background. Currently, modeling species with restrict datasets assumes particular importance, once the ongoing global changes in climate and land cover may not allow waiting for more comprehensive species databases. Although there are techniques able to overcome the difficulties of estimating species distributions under small datasets, much less is known about modeling approaches able to accurately predict outside the spatial range of a small sample.

Methods. We compared the performance of Maxent and the Bayesian technique 'Weights of Evidence' regarding their ability to maintain satisfactory suitability surfaces even when the data are restricted, both in terms of quantity and location. In order to achieve this goal, the modeling methods were applied to three Neotropical ant species, and for each of these, the models were run using three distinct datasets: (a) the complete dataset consisting of species presence data, (b) a restricted dataset – both in terms of number of records and geographical coverage of the data – also based on species presence data, and (c) a restricted dataset whose records consisted of abundance rather than presence data. Evaluation of results were conducted with independent presences and absences through comparisons of True Skill Statistics (TSS), a metric that ranges from 1 (complete agreement) to -1 (complete disagreement).

Results. Concerning modeling with complete datasets, TSS values estimated in a 23 x 23 km window size varied from 0.5 to 1 with Weights of Evidence, and from 0 to 0.31 with Maxent. When calibration data were restricted, Maxent had their TSS estimated in a 23 km x 23 km window varying from 0.12 to 0.76, whereas Weights of Evidence showed TSS values higher than 0.72 in this same window size. Overall, Weights of Evidence outperformed Maxent because it was able to predict many of the presence records located outside the geographical limits of restricted datasets and also because it was more accurate in predicting absences. The inclusion of abundance values in modeling restrict datasets did not significantly improve TSS.

Discussion. Our results suggest that Weights of Evidence is more robust than Maxent, especially when dealing with restrict databases. Given this circumstance, only Weights of Evidence was able to classify correctly almost all absences and presences located outside the limits of the calibration ones. We believe Weights of Evidence performed better than Maxent mainly as a result of the calibration strategy applied here. We adopted

a new approach for running Weights of Evidence, so that the number and location of pseudo-absences during the calibration process were specifically designed for each modeled scenario. To our knowledge, this is the most comprehensive study to evaluate species distributions obtained through Weights of Evidence, a method traditionally used in geosciences. We thus encourage further studies regarding its application in species distribution modeling.

Introduction

Modeling species spatial distribution is an important tool in conservation planning (Wilson et al., 2005). However, its application has been especially difficult for species that have only few occurrences, mainly because these may be covering only a small portion of the real geographic range of species. Over the past few years, there has been an ongoing effort to develop techniques designed to overcome the challenges of modeling under small datasets (Wisz et al., 2008; Hernandez et al., 2006; Breiner et al., 2015; Hamilton et al, 2015; Galante et al., 2017). Notwithstanding, little is known about modeling approaches able to accurately predict outside the spatial range of a small sample. This issue is of special concern to develop conservation strategies for the tropics. Despite being the region with the highest biodiversity, there is a shortage of information regarding species occurrences (El-Gabbas et al., 2017). The rapid changes in climate (Magrin et al., 2014) and land cover (Ferreira et al., 2013; Soares-Filho et al., 2006) may not allow waiting to collect more comprehensive species datasets.

Recent assessments revealed that, the most reliable approach for modeling species with few occurrence records (< 30) is to apply the Maxent algorithm (Philips et al., 2006; Hernadez et al., 2006; Wisz et al., 2008) using the ENMeval package (Muscarella et al. 2014; Galante et al., 2017). ENMeval (Muscarella et al. 2014) is an R (R Core Team, 2017) package that builds a series of Maxent models with a variety of settings in order to select the optimal model for a given species under few occurrences. Bayesian networks and hierarchical models are other important modeling techniques that have been successfully applied to small datasets (Wilson et al., 2010; Hamilton et al., 2014). They allow incorporating sources of uncertainties, such as indirect signs of the species (Wilson et al., 2010), bias correction (Velásquez-Tibata et al., 2015), and *à priori* probabilities stated by species' experts (Hamilton et al. 2015) and, hence, they minimize the errors associated with modeling under few data. Nonetheless, in contrast to Maxent, these approaches are far from simple. Their applications involve both complex theoretical

considerations and intensive computation. In addition, they require additional information that are not available for the majority of species with few records.

Nevertheless, an important uncertainty still remains regarding modeling methods: it is not known which modeling technique has the greatest performance in predicting outside the limits of a geographically restricted dataset. This kind of analysis - frequently called as transferability assessment - is gainning attention (Wenger et al., 2012, Breiner et al., 2015) due to its importance in conservation planning. If a model has a satisfactory performance only around the input presence points and cannot detect other suitable areas outside the current known species range, it means that it is not able to predict new potential areas for the species. The ability of indicating new suitable areas for a given species is crucial, especially because of climate change, which is expected to alter species current geographical ranges (Chen et al., 2011). To our knowledge, only five studies have evaluated this issue (Wenger et al., 2012; Breiner et al., 2015; Muscarella et al., 2014, Qiao et al., 2018; Iturbide et al., 2018), with only one considering restricted datasets (Muscarella et al., 2014), and none of them comparing modeling algorithms.

Within this context, we compared the performance of a Bayesian probability method, Weights of Evidence (Agtberg and Bonhan-Carter, 1990), with Maxent (Philips et al., 2006) regarding both their ability to deal with restricted datasets and their potential to be used for transferability purposes, using a non-biased validation technique. We conducted modeling experiments by varying both the number of occurrences and their spatial ranges for three species of ants with different number and spatial patterns of presence/absence data. In order to avoid validation bias, evaluation procedures were conducted with independent - hold-out data - by analyzing TSS (True Skill Statistics), sensitivity and specificity indexes in different scales together with visual inspections of modeled probability maps. This is the first paper to compare Weights of Evidence to Maxent in modeling species with geographically restricted data.

Materials & Methods

We compared Maxent and Weights of Evidence regarding their ability to maintain satisfactory suitability surfaces even when the data are restricted, both in terms of quantity and location. In order to achieve this goal, each modeling algorithm was applied to three species and for each species, the models were run using three distinct datasets: (a) the

complete dataset consisting of species presence data, (b) a restricted dataset – both in terms of number of records and geographical coverage of the data – also based on species presence data, and (c) restricted dataset whose records consisted of abundance rather than presence data. The combination of dataset type and modeling algorithm resulted in six scenarios for each species. These scenarios were then evaluated concerning validation scores (based on independent test data). In the following sections, each of these steps are described in details.

Data

As explaining variables, we considered factors that might shape species distributions: climate, the amount of energy available in the system, and the topographic characteristics of the terrain. In this study, we obtained climatic conditions from WorldClim database, version 1.4 (Hijmans, 2005), constituting of 19 maps representing combination of means, extremes, and seasonality in rainfall and temperature. To characterize system energy, we used net primary productivity and gross primary productivity. Both are available from MODIS sensor (https://lpdaac.usgs.gov/products/modis_products_table/mod17a3). Altitude and slope were derived from SRTM (Shuttle Radar Topography Mission) maps. All maps represent current conditions in 1 km resolution and span South America.

The restricted datasets were derived from the study of Vasconcelos et al. (2017). In this study, a systematic sampling of the arboreal and ground-dwelling ant faunas was conducted in 29 sites spread haphazardly over a region of approximately 2000 x 1800 km in the Brazilian *Cerrado* (savanna) biome (Fig. 1). In each site, it was stablished three sampling transects (≥ 1 km apart from each other), with a total of 20 sampling plots each (spaced 20 m from each other). Thus, in each site there was a total of 60 sampling points and ant abundance data represented the number of samples (for a total maximum of 60) in which a given species was recorded. For our study, we selected three of the species collected by Vasconcelos et al. (2017). These were: *Cyatta abscondita, Gracilidris pombero*, and *Linepithema cerradense*. These species were selected because they showed varying patterns of distribution and abundance. *Linepithema cerradense* was both widely distributed across the *Cerrado* biome and locally abundant (it was present in 27 of the 29 sites sampled and its local abundance varied from 2 to 31; mean = 16 species records per sampling site). *Gracilidris pombero* was present in 13 sites and its abundance varied from 1 to 13 (mean = 6). *Cyatta abscondita* was the rarest species, found in only five sites, at

an abundance ranging from 1 to 3 species records per site. Spatial coordinates of presences corresponded to either the location of the single transect in which the species was recorded or the centroid of the transects, in case the species was present in more than one transect. Sites in which a given species was not collected were considered as true absences, and spatial location was also attributed to transects centroid. Although restriction condition is *a priori* limited to the *Cerrado* boundaries - due to the sampling design of Vasconcelos's research - *Gracilidris pombero* and *Cyatta abscondita* showed an even smaller geographic range under restriction condition, given their small number of occurrences (Fig 1).

The "complete" dataset consisted of both the data described above and data obtained from antmaps.org (Janicki et al., 2016). The antmaps.org data (www.antmaps.org) is a comprehensive global database of ant species distributional records, including literature records, museum databases and online specimen databases, that nevertheless has not yet included the data collected by Vasconcelos et al. (2018). The data downloaded from antmaps.org was scanned for reducing possible bias, retaining only species occurrences whose sources have been published in the literature and removing records that were less than 2 km apart from each other, as these could inflate predictions. The total number of occurrences (antmaps data plus Vasconcelos et al., 2018 data) then comprised 44 for *L. cerradense*, 31 for *G. pombero*, and 16 for *C. abscondita*.

Species data were divided into calibration (or training) and validation (or testing) subsets (Fig 1). In the modeling scenarios that used the complete datasets, validation data represented 25% of the presence occurrences, randomly chosen through systematic sampling ordered by latitude values of presences, plus the absences. For models using restrict datasets, validation subsets for *G. pombero* and *C. abscondita* were composed of all occurrences from antmaps and all absences recorded by Vasconcelos (2017). However, for *L. cerradense*, the validation subset consisted of all occurrences from antmaps presences and eight randomly selected occurrences from Vasconcelos et al. (2017). This was done in order to create a significant difference in number of calibration occurrences between the complete and restrict modeling experiments for this species. Otherwise, calibration values for complete and restrict datasets would be very similar (33 in the complete dataset and 27 in the restrict one). The range of calibration occurrences in complete and restrict datasets led to distinct areas, as pointed out by the minimum

boundary polygon surrounding calibration points (Fig. 1), allowing us to perform transferability assessments.

Modelling methods

For each of the nine species scenarios, Maxent suitability surfaces were obtained through two steps. Firstly, we applied the ENMevaluate function (Muscarella et al., 2014) to calibration data using the jackknife option. As a result, this function returns a table with several parameters and statistics, among them, Feature Class, Regularization multiplier, AIC and delta AIC. The combination of parameters Feature Class and Regularization Multiplier that resulted in the best fitted model, as demonstrated by a delta AIC equal to zero, was used to run the final model of each one of the nine species dataset scenarios in Maxent software, version 3.4.1. In this software, the final suitability surfaces were obtained by averaging ten subsampled models, each run with seventy percent of calibration data. In order to turn Maxent surfaces of suitability values into probability ones, the clog-log transform was applied (Philips et al., 2017).

While Maxent and the ENMevaluate function could be easily applied to the data because they were specifically designed to the case of species presence information, the use of Weights of Evidence required several modifications. In the next sections, we briefly describe the Weights of Evidence and, following, we detail the steps we developed to fit this technique to the case of modeling species distributions with small datasets of presence-only data.

Weights of evidence

Weights of Evidence (hereafter referred to as WE) is a Bayesian method traditionally used to derive suitability maps for spatial point phenomena (Agterberg and Bonham-Carter, 1990; Bonham-Carter, 1994; Soares-Filho et al., 2009). We believed in its potential to model species distribution due to the fact that it was originally developed to model mineral occurrences whose input dataset constituted solely on drilling holes located across the geographical space, with no information on sites in which these holes were not taken. Therefore, being an analogous situation as modelling species distributions with presence only data.

WE consists on analyzing a set of explanatory variables (in raster map format) and to associate them with known occurrences of the event. These associations are estimated for several categories of each one of the explanatory maps by means of odds ratios of

conditional probabilities, based on Bayes rule, and are called as Weights of Evidence (Bonham-Carter, 1994). The magnitude of the weights depends on the measured association between the phenomena and the explanatory variables: positive values for Weights of Evidence favor the modeled event, the higher the value, the stronger the association. On the other hand, a negative weight indicates an inhibitory effect, whereas values close to zero are consistent with no association (Bonham-Carter, 1994). The probability of occurrence of the phenomenon in a given raster cell is estimated by integrating all variable weights of that cell through posterior logits, as demonstrated in Eq. 1. For more details concerning derivation of weights and the probability equation, the reader should refer to Bonham-Carter (1994).

$$P(A \cap B \cap C.. \cap N)_{(x,y)} = \frac{e^{\sum_{i=A}^{N} W_i}}{1+e^{\sum_{i=A}^{N} W_i}} \qquad \text{Eq. 1}$$

where,

$P(A \cap B \cap C.. \cap N)_{(x,y)}$ is the probability of occurrence of the phenomenon at the (x,y) cell given N explaining variables (A, B, C, … N), whereas W_i is the weight of variable i.

The only assumption of this method is that all explanatory variables must be independent. The Weights of Evidence is implemented in Dinamica EGO modelling platform (Soares-Filho et al., 2009) and only requires raster maps of explaining variables and occurrence data to be applied. By default, information of the entire geographic range of the input maps are used in this process, assuming that the whole area was sampled. Given this assumption is not true for the species databases used here, we applied a series of modifications in the way WE is traditionally calibrated (see Soares-Filho et al., 2009). The fitting process used here will be described in details bellow and was applied to each species scenario.

1. Generating input data

Instead of using maps of the entire South America, we used just a sub-sample of them. This was required in order to adequate the modeling algorithm into the case study of modeling species under data-poor conditions, in which the real geographic range of the species may have not been entirely sampled.

Firstly, as the geographic domain influences the location of pseudo-absences, which, in turn, can have significant effects on model performance (VanDerWall et al., 2009; Wisz and Guisan, 2009; Lobo et al., 2010), we defined calibration regions as the minimum boundary polygon surrounding presence points (Fig. 1) added to its 500 km buffer. The idea behind this approach is to create pseudo-absences that are not too close to presence records, what would cause spurious predictions (VanDerWall et al., 2009) and also to avoid inflating validation statistics, caused by over spreading of pseudo-absences (Lobo et al. 2010). Other ranges were tested but 500 km produced good results for all the species. Then, a randomly selected subset of the background cells composed the pseudo-absences group. Number of pseudo-absences were determined as the sum of presence weights, as suggested by Barbet-Massin (2012), who showed that models fitted with a large number of pseudo-absences, but equally weighted to presences, produces the most accurate predicted distributions. After testing several approaches to estimate the best values to be attributed to presences, we established the following rules: 1) for complete datasets and restrict ones without the use of abundance data, each presence occurrence is given a weight of 10; 2) for the restrict dataset, in which abundance values were taken into account, presence weights corresponded to observed abundance for *L. cerradense* and to the observed abundance (times 10) for *G. pombero* and *C. abscondita*. We decided to multiply abundance by 10 for these two species to increase the importance of the presence occurrences in relation to pseudo-absences in modeling procedures, since *G. pombero* and *C. abscondita* showed low abundance values.

Then, the weights of the presences, the pseudo-absences and explaining variables were organized into a table whose first column indicated presence or pseudo-absence - assuming the presence weight or the value of zero, in case of pseudo-absences - and the next columns showed the values of explanatory variables collected at the presences and pseudo-absences points. Following, each column was organized as a matrix, hereafter called fiction raster maps, which were used as inputs in Dinamica EGO WE calibration process, instead of using maps of entire South America. The name "fiction" maps was chosen because they are, actually, simple matrices, whose cells do not have any geographical relationships with one-another. Fiction maps of species and explanatory variables are structured in the same way and have the same number of lines and columns, and used in asc format. Although these maps do not have any spatial relation between

cells, this fact does not have any implication on modeling estimations, since the WE does not depend on any kind of neighborhood relationships.

The process of creating the fiction rasters was developed in R (R Core Team, 2017) and the code is available for download. To be run, the function requires the geographical coordinates of presences and a map showing the contiguous area in which the presence points are inserted (in our case, it was South America). By default, it assumes presence weights will be ten and it will generate a number of pseudo-abscences correspondent to number of presences times 10. For other weights, the user should alter the source code.

2. Calculating Weights of Evidence

The fiction maps were passed on to Dinamica EGO functions "Determine weights of evidence ranges" and "Determine Weights of Evidence Continuous occurrence" (Agterberg and Bonham-Carter 1990; Soares-Filho et al., 2009), which were run iteratively, changing the parameters used to estimate the ranges at each step. Interactions were obtained by varying the following parameters of the "Determine weights of evidence function": minimum delta (assumed values 1, 20 and 30), tolerance angle (could be 5 or 10) and increment (varying between 1 and 20% of variable amplitude). After each step, the significant weights and their ranges were saved in a separated txt file. Iterations were run as many times as necessary in order to fine-tune and reach the broadest sequence of significant range weights for each variable. Given overlaps between ranges originated from different iterations, we preferred the longer ones and the ones with higher contrast (contrast is one of the parameters that Dinamica Ego uses to evaluate weight coefficient significance). In sum, the main goal of these iterations was to seek for variables whose weight's graph presented a clear trend, easy to interpret and to establish relations between the explanatory variables and species occurrences. Given a graph whose general trend is clear, but that presents one or two classes whose weights fall far away from their trend or is equal to zero, these classes' weights were assigned to the mean of its neighbors' coefficients, as suggested by Soares-Filho et al. (2009). This is an approach to correct values that might have been occurred due to chance, by a non-sampled place, for example.

Usually, one certain combination of these parameters is enough to generate significant weights (Soares-Filho et al., 2009) and a weights' graph showing a clear trend. However, in the case of modelling species with few occurrences, as the one used here, we noted that a unique combination of parameters was never enough to generate ranges

with significant weights spanning across the amplitude of a certain variable. This might have happened due to a limited ability of the method used to categorize the continuous variables given a small dataset.

Within this step we applied the first filter used to select the most important variables related to species occurrence: only explanatory variables with more than 70% of significant classes are considered in the next step.

3. Estimating probability of occurrence given just one variable

From this step on, the fiction maps are no longer used and we estimated probability maps of species occurrence given just one variable along the entire South America. As ranges and weights were estimated based on a buffer polygon surrounding presences, areas of South America outside the polygon remained without ranges and weights. Hence, two new ranges were created in the upper and lower limits of the sequence of ranges. Given a certain variable, the lower range varies from its minimum value observed in the South America to the minimum value in its fiction map, and the upper range begins on the maximum value of the fiction map and closes on maximum value of the variable observed in the South America map. Weights assigned to the lower and upper ranges assumed a value of 20% less than the most negative weight obtained during calibration or, in case of all calibrated weights being positive, weights of upper and lower classes were assigned to -0.5.

Each variable had their weights applied to its South America map using the function "Calculate Weights of Evidence Probability map" in Dinamica EGO (Soares-Filho et al., 2009). The idea behind this approach is to evaluate how much of the species distribution can be explained alone by each variable.

In this phase, we applied two more filters in order to select important variables. To be considered in the next steps, variables should have sensitivity higher than 75% and partial Area Under Curve (Peterson & Soberon, 2008) higher than 80%. Both these statistics took into account calibration presences.

4. Checking correlation between variables

Correlation between pairs of selected variables were evaluated through the Joint Information Uncertainty Index, estimated by the Dinamica EGO function "Determine Weights of Evidence Correlation" (Bonhan-Carter, 1994; Soares Filho et al, 2009). The theoretical limit of this correlation coefficient varies from 0 to 1 and, in our analysis, for

a given pair of variables with correlation higher than 0.5, just the one with the higher sum sensitivity plus partial ROC remained in the model.

     5.   Estimating species distributions

We obtained current and future probability maps of species occurrence by applying the function "Calculate Weights of Evidence Probability map" into selected variables. This function applies equation 1 into explaining variables. Then, suitable and non-suitable areas were defined based on a probability threshold that maximizes the sum sensitivity plus specificity. This threshold was obtained using the function "Find.Optim.Stat" in biomod2 package (Thuiller et al., 2014) within R (R Core Team, 2017), using ROC as the parameter.

Evaluation of model predictions

We validated modeled predictions by evaluating True Skill Statistics (TSS, Allouche et al., 2006) across different scales and, at the same time, associated them to percentages of correct classification of presences (sensitivity) and absences (specificity) separately. In parallel, we visually analyzed the maps of predicted probability.

Firstly, TSS (Alouchee et al., 2006) was applied to compare predicted and observed occurrence maps of presences and absences. To obtain predicted maps, we used the threshold value of 0.5: cells with probabilities greater than 0.5 were considered as presences, otherwise they were classified as absences. To analyze TSS in different scales, we estimated percentages of correct classification on window sizes varying from 1 to 23 cells, corresponding to a spatial resolution ranging from 1 to 23 km. This analysis was also applied for both presences and absences. While TSS will give a general view of agreement between observed and predicted maps, the sensitivity and specificity will show how well the model predicts these kinds of occurrences separately. We run this approach in Dinamica EGO (Soares-Filho et al., 2009) by applying the "Calculate Reciprocal Similarity Map" into maps of validation presences and absences.

Probability maps were visually evaluated in parallel to these indexes. We believe that in spatial modeling, visual inspections of resultant probability maps assume the same importance as the analysis of summary statistics, since they can bring into light patterns that could not be revealed by a mathematical analysis. This is especially important in the case of species modeling, where, in general, data are not systematically collected along the entire region. In this context, the aim was to evaluate not only if the predicted surface

is in accordance with the pattern of the species database but also to see how the model behaves in areas where there is no information about the modeled species. Moreover, with restricted datasets, when one omits a careful visual inspection and uses only metrics instead, evaluation of model predictions becomes more susceptible to misleading conclusions regarding predicted accuracy. For example, if a suitability map presents higher values only close to presence records, TSS, sensitivity and specificity will return satisfactory values. Nonetheless, predicted suitability would be poor, bringing no knowledge gain in relation to what is already known from the collected data.

Results

The general results were consistent among the three species analyzed, with Weights of Evidence showing better predictions than Maxent for all datasets conditions (Fig. 2).

Considering modeling of complete datasets, validation TSS lines for Weights of Evidence were higher than those from Maxent for all window sizes and species. By analyzing *G. pombero* in a cell by cell comparison, while Maxent TSS corresponded to -0.25, Weights of Evidence was equal to 0.19 (Fig. 2A), a difference of 0.44. With the increase of window size to 23 x 23 km, TSS increased to 0.31 in Maxent and to 0.5 in WE (Fig. 2A), reducing the difference in accuracy rates to only 0.19. For *L. cerradense*, although TSS of WE and Maxent were the same in a 1x1 km comparison, at the 23x23 km window size Maxent TSS was equivalent to 0 while WE TSS corresponded to 1 (Fig. 2D). Maxent TSS for *C. abscondita* in a cell by cell comparison was 0.26 (trocar por 28%) points lower than the value that resulted from the WE (Fig. 2G). Considering the largest window size analyzed, this difference increased to 0.46 (trocar por 52) (Fig. 2G). The main cause of these observed differences was prediction of absences. Although accuracy in predicting presences were high for all species and for both modeling algorithms (equal to 100% - except for *G. pombero* modeled by Maxent which was equivalent to 87.5% - Figs. 2B, 2Ee 2H), Maxent was less accurate in predicting absences than WE (Figs. 2C, 2F e 2I). While WE percentages of correct classification for absences in the largest resolution analyzed corresponded to 100% for *L. cerradense*, 71(trocar por 80%) % for *C. abscondita* and 50% for *G. pombero*, these same values for Maxent were 0%, 25% and 44%. The analysis of probability maps (Fig 3A-B, 4A-B and 5A-B) shows that the outlines that can be made by gathering the highest probability areas in WE

resulted in spatial shapes with irregular contours, on the opposite of Maxent ones, whose highest probability regions presented more smooth borders and hence, more homogenous polygons.

The analysis of restrict datasets showed that WE produced the best-fit models regardless the use of abundance values. WE TSS of restrict validation datasets almost reached total accuracy in a 23 x 23 window size for *G. pombero* and *L. cerradense* and corresponded to 0.65 for *C. abscondita*, which was calibrated with only 5 presence occurrence points (Figs 2A, 2D, 2G). When comparing *C. abscondita* probability maps originated from restrict and complete datasets using WE, we found that even the map from the restricted dataset shows a huge contraction of the highest probability area (Fig. 2I), it still predicted the majority of validation points correctly (90% of absences and 80% of presences in a 23 x 23 km window size). The probability map of *L. cerradense* for both Maxent and WE showed an increase in the highest probability area in relation to the map obtained with the complete dataset, incorporating the Andes as a suitable region for the species. In contrast to WE TSS of restrict datasets, Maxent TSS reached a maximum of 0.76 for *G. pombero*, 0.38 for *L. cerradense* and 0.12 for *C. abscondita* in a 23 x 23 km window size (Fig. 2A, 2D, 2G). The decrease in performance rates from WE in relation to Maxent was due to the more general probability surfaces resultant from Maxent (Figs. 3C-D, 4C-D, 5C-D), making specificity rates of this method much lower than those from WE. This effect was more pronounced for models involving *C. abscondita*, which showed a too general probability surface.

The inclusion of abundance values on modeling restrict datasets caused different effects on the two modeling techniques and, overall, did not significantly improve the TSS statistics. In general, while Maxent surfaces have their highest probability regions reduced in size with the inclusion of abundance values for *G. pombero* and *L. cerradense*, WE maintained similar areas of highest suitability but increased probability values of these areas for the three species (Figs. 3E-F, 4E-F, 5E-F). The small differences in Weights of Evidence TSS rates caused by including abundance in restrict datasets were not enough to suggest that abundance values could improve predictions of this method (2A, 2D, 2G). However, we noted one positive modification for *L. cerradense*: the use of abundance values caused a decrease probability of occurrence in the Andes (Fig 4C-F). Regarding Maxent predictions including abundance information, *C. abscondita* showed a huge expansion of the highest probability area (Fig. 5E), maintaining sensitivity values

equal to 100% but reducing even more specificity, downgrading to 0% (Figs. 2H and 2I). Maxent TSS of restrict with abundance datasets showed higher values than the restricted scenarios for *G. pombero* and *L. cerradense* (Figs. 2A, 2D, 2G). Nevertheless, a closer analysis supported by both the probability maps (Figs. 3F and 4F) and sensitivity and specificity graphs (Figs. 2C and2F), shows that they have predicted really few presences, a result of the contraction observed in the highest probability area (Figs. 3F and 4F).

Discussion

Our results suggest that, in general, Weights of Evidence (WE) performed better than Maxent for all datasets conditions for the three species analyzed. Concerning modeling with complete datasets, while TSS estimated in a 23 x 23 km window size varied from 0.53 to 1 for WE, Maxent values ranged from 0 to 0.34. When calibration data were restricted to *Cerrado* boundaries, WE TSS were higher than 0.72 for all species and Maxent values ranged from 0.12 to 0.75. Taking into account that TSS varies from -1 to 1, these observed differences in TSS among modeling algorithms suggest that Maxent predictions are significantly less accurate than those from WE for the complete datasets. The reason why Maxent performed poorly than WE was due to its low accuracy for absences in both datasets and its poor ability in predicting validation presences located far away from calibration region. These results, therefore, suggest that WE might be transferable to unsampled regions nearby calibration points with a superior accuracy than Maxent. WE most outperformed Maxent for the species with the smallest dataset, *C. abscondita*, for which Maxent returned too general probability surfaces – showing high values in a large contiguous area of the map, which, hence, does not allow absences to be predicted. The really extensive suitable area of *C. abscondita* (Figure 5D) is the opposite of the expected for a species with a restrict distribution.

It was interesting to note that WE TSS of restrict datasets for *G. pombero* and *L. cerradense* were even higher than WE TSS of complete datasets for window sizes greater than 5x5 km. We believe this might be an effect of the lower variance presented by the explanatory variables of the restrict dataset presence points in relation to the complete dataset occurrences. As restrict datasets were collected only in the *Cerrado* biome, the climatic and environmental characteristics of these points show a smaller variability than the ones that were collected in several other biomes. This low variability resulted in a narrower range of suitable environmental conditions than the one obtained with complete datasets, resulting in a straight forward search for environmental conditions similar to the

ones in which the species were found within *Cerrado*. The inclusion of abundance values caused little differences on WE modeling which, overall, were not strong enough to suggest that they improved predictions. For Maxent, visual analysis of probability maps suggests that the use of abundance data may even have a negative impact on model predictions.

The basic of statistics states that the liability to trust a model's output is directly linked to three main factors: 1) the quality of the input data; 2) the evaluation strategy applied; and 3) the degree to which the assumptions of the method are met. Here, we attempted to control all these three factors, since lack of attachment to these rules may lead to greater uncertainty in predictions. The input species data used here were either obtained from a planned survey across a large region (the Brazilian *Cerrado*), and thus represented a relatively spatially unbiased dataset, or from presence records gathered from different sources by antmaps.org, which were filtered in order to remove samples too close to one another and that, therefore, could inflate predictions. Moreover, the spatial error that might have been included due to the use of explanatory variables obtained from remote sensing imagery was diminished by applying spatial filters on predicted suitability maps.

Considering the evaluation strategy, a modeler should attain to adopt the one(s) with best fit to their dataset, mainly concerning presence-only or presence-absence data availability. We chose to rely on TSS (Alouchee et al, 2006), a non-biased technique which has a really straight-forward meaning, is robust and easy to interpret. Along with this metric, we evaluated percentages of correct classifications of presences and absences, and conducted visual analysis of predicted probability maps. Given the lack of a planned survey across the entire region being modeled, the visual analysis of probability maps assumes particular importance. In this work, *L. cerradense* serves as an example regarding this issue. Under restrict datasets, although both modeling methods showed satisfactory evaluation metrics, their probability surfaces predicted the Andes region as a suitable area for the species. However, as far as we know, there are no *L. cerradense* records in the eastern side of the Andes, and it certainly does not occur in the western side since the Andes itself acts as a barrier to the dispersal of *L. cerradense*.

Regarding modeling methods assumptions, our approach relied on one of the types of models more appropriate to spatial datasets, the Bayesian methods (Bonham-Carter, 1994). It is well known that Bayesian methods are adequate for spatial information since

they are estimated in a nonparametric way, and thus not presenting all the restrictive assumptions of parametric statistical models. In comparison to Hierarchical Bayesian techniques that have been increasingly applied to species distribution modelling (Wilson et al., 2010; Carrol et al., 2008, Rivera et al., 2008), Weights of Evidence is a simpler method regarding equations and methodologies used to estimate parameters. Nevertheless, although it is already implemented in a user-friendly modeling software (Dinamica EGO, Soares-Filho et al., 2009), the WE fitting process applied here was time consuming for two reasons: 1) we created fiction raster maps to be used in the calibration process, constituting matrices showing presences and pseudo-absences as explanatory variables; and 2) the fine tuning of Weights of Evidence categorization process required exhaustive modification in the parameters in order to reach the categorization pattern with the greatest number of significant classes. Futures researches could include the automatization of these process, allowing WE to be readily applied to model species with few and spatially restricted occurrences.

A recent work comparing Weights of Evidence to Maxent suggested that these methods have similar efficacy levels (Lopes et al., 2014). We believe we reached a distinct conclusion both because we used a different evaluation strategy and also due to the different way of conducting the calibration process. While Lopes et al (2014) used AUC statistics taking into account only presence records, we applied a non-biased metric, TSS, using both presences and absences. In addition, our calibration process was based on the creation of fiction maps that turned WE more appropriate to the case of species distribution modeling. These maps allowed us to control the number of pseudo-absences and their weights in relation to each presence record, as suggested by Barbet-Massin (2012). In addition, they made it possible to restrict these pseudo-absences to buffer zones related to the extent of calibration presences. Restricting model fitting to an area surrounding occurrence points is a key factor to obtain more realistic predictions than models calibrated within areas that extrapolates species occurrence ranges (Anderson & Raza, 2010).

Conclusions

Overall, it was worthwhile to conduct species distribution modeling through the Weights of Evidence method, despite the time-consuming calibration process adopted here. Our results suggest that WE is more robust than Maxent, especially when dealing with restrict databases, in which case only WE was able to classify correctly almost all

presences and absences located outside the limits of the calibration ones. This is an imperative behavior of models to be applied in datasets whose range do not span the entire geographic distribution of the species and an especially important ability when the aim of modeling is to design conservation strategies in face of climate and land use change (Peterson et al., 2007; Qiao et al., 2018; Iturbide, Bedia & Gutierrez, 2018). We highlight the need of further investigations regarding the use of Weights of Evidence in species distribution modeling, including the use of different species and a more refined transferability assessment, considering different levels of split sampling.

References

Agterberg FP, Bonham-Carter GF (1990) Deriving weights of evidence from geoscience contour maps for the prediction of discrete events. Pages 381–395 in XXII International Symposium. Volume 2. APCOM, Berlin,Germany.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied Ecology, 43: 1223–1232.

Anderson RP, Raza A (2010) The effect of the extent of the study region on GIS models of species geographic distributions and estimates of niche evolution: preliminary tests with montane rodents (genus Nephelomys) in Venezuela. Journal of Biogeography, 37(7), 1378-1393.

Araújo MB, Guisan A (2006) Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33, 1677–1688.

Barbet-Massin M, Jiguet F, Elbert CH, Thuiller W (2012) Selecting pseudoabsences for species distribution models: how, where and how many? Methods in Ecology and Evolution 3, 327–338

Bonham-Carter G (1994) Geographic information systems for geoscientists: modelling with GIS. Pergamon, Oxford, UK.

Breiner FT, Guisan A, Bergamini A, Nobis MP (2015) Overcoming limitations of modelling rare species by using ensembles of small models. Methods in Ecology and Evolution 6, 1210–1218.

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociological Methods and Research 33:261–304

Carroll C, Johnson DS (2008) The importance of being spatial (and reserved): assessing northern Spotted Owl habitat relationships with hierarchical Bayesian models. Conservation Biology, 22, 1026–1036.

Chen IC, Hill JK, Ohlemuller R, Roy DB, Thomas CD (2011) Rapid range shifts of species associated with high levels of climate warming. Science 333, 1024–1026.

El-Gabbas A, Dormann CF (2017) Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. Ecography, https://doi.org/10.1111/ecog.03149

Ferreira ME, Ferreira LG, Miziara F, Soares-Filho BS (2012) Modeling landscape dynamics in the central Brazilian savanna biome: future scenarios and perspectives for conservation. Journal of Land Use Science 8, 403-421.

Galante PJ, Alade B, Muscarella R, Jansa SA, Goodman SM, Anderson RP (2017) The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. Ecography. https://doi.org/10.1111/ecog.02909.

Hamilton SH, Pollino CA, Jakeman AJ (2015) Habitat suitability modelling of rare species using Bayesian networks: Model evaluation under limited data, Ecological Modelling, 299, 64-78.

Hernández PA, Graham CH, Master LL, Albert DL (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29, 773–785.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. International journal of climatology, 25(15): 1965-1978.

Holland JH (1975) Adaptation in Natural and Artificial Systems (Ann Arbor: University of Michigan Press).

IPCC, 2007: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M.

Tignor, and H. L. Miller (eds.). Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 996 pp. (2007).

Iturbide M, Bedia J, Gutierrez JM (2018) Background sampling and transferability of species distribution model ensembles under climate change, Global and Planetary Change, doi:10.1016/j.gloplacha.2018.03.008

Janicki JH, Narula N, Ziegler M, Guénard B, Economo EP (2016) Visualizing and interacting with large-volume biodiversity data using client-server web mapping applications: the design and implementation of antmaps.org. Ecological Informatics, 32: 185-193.

Jaynes ET (1957) Information theory and statistical mechanics. Physical Review, 106, 620–630.

Lobo JM, Jiménez-Valverde A, Real R (2008) AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17, 145–151.

Lobo JM, Jimenez-Valverde A, Hortal J (2010) The uncertain nature of absences and their importance in species distribution modelling. Ecography 33:103–114

Lopes DFMT, Silveira AC, Soares-Filho BS (2014). Evaluating Weights of Evidence method for habitat suitability modeling: a comparison with Maximum Entropy for a case of few species records. Available at: https://www.csr.ufmg.br/dinamica/dokuwiki/doku.php?id=habitat_suitability_modeling_with_small_sample_size

Magrin GO, Marengo JA, Boulanger JP et al (2014) Central and South America. In: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Barros, V.R., C.B. Field, D.J. Dokken, M.D. Mastrandrea, K.J. Mach, T.E. Bilir, M. Chatterjee, K.L. Ebi, Y.O. Estrada, R.C. Genova, B. Girma, E.S. Kissel, A.N. Levy, S. MacCracken, P.R.

Mastrandrea, and L.L. White (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, pp. 1499-1566.

Marini MA, Barbet-Massin M, Lopes LE, Jiguet F (2009) Major current and future gaps of Brazilian reserves to protect Neotropical savanna birds. Biological onservation, 142:3039–3050.

Muñoz MES, Giovanni R, Siqueira MF, Sutton T, Brewer P, Pereira RS, Canhos DAL, Canhos VP (2009) OpenModeller: a generic approach to species' potencial distribution modeling. GeoInformatica, 1-25.

Muscarella R, Galante PJ, Soley-Guardia, M, Boria RA, Kass JM, Uriarte M, Anderson RP (2014) ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods in Ecology and Evolution, 5, 1198–1205.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190, 231-259.

Phillips SJ, Dudík M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31, 161–175.

Phillips SJ, Anderson RP, Dudík M, Schapire RE, Blair ME (2017) Opening the black box: an open-source release of Maxent. Ecography. http://dx.doi.org/ 10.1111/ecog.03049

Qiao H, Feng X, Escobar LE, Peterson AT, Soberon J, Gengping G, Papes M (2018) An evaluation of transferability of ecological niche models. Ecography, DOI: 10.1111/ecog.03986

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Radosavljevic A, Anderson RP (2013) Making better Maxent models of species distributions: complexity, overfitting, and evaluation. Journal of Biogeography, 2014, 41, 629-643.

Rivera OL, Lópes-Quílez A, Blangiardo M (2018) Assessing the Spatial and Spatio-Temporal Distribution of Forest Species via Bayesian Hierarchical Modeling. Forests 2018, 9, 573; doi:10.3390/f9090573

Shcheglovitova M, Anderson RP (2013) Estimating optimal complexity for ecological niche models: a jackknife approach for species with small sample sizes. Ecological Modelling 269, 9–17.

Soares-Filho, BS, Nepstad DC, Curran LM et al. (2006). Modelling conservation in the Amazon basin. Nature 440:520-523.

Soares-Filho BS, Rodrigues HO, Costa WL (2009). Modeling Environmental Dynamics with Dinamica EGO. https://www.csr.ufmg.br/dinamica/dokuwiki/doku.php?id=tutorial:start. (Acessed May, 2018).

Stockwell DRB, Noble IR (1992) Induction of sets of rules from animal distribution data: a robust and informative method of data analysis. Mathematics and Computers in Simulation 33, 385–390.

Stockwell D, Peters D (1999) The GARP modelling system: problems and solutions to automated spatial prediction. International Journal Geographical Information Science 13, 143–158.

Thuiller W, Georges D, Engler R (2014). biomod2: Ensemble platform for species distribution modeling. R package version 3.1-64. http://CRAN.Rproject.org/package=biomod2

VanDerWal J, Shoo LP, Graham C, Williams SE (2009). Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? Ecological Modelling, 220:589–594.

Velásquez-Tibatá J, Graham CH, Munch SB (2015) Using measurement error models to account for georeferencing error in species distribution models. Ecography, 38, 001–012.

Vasconcelos HL, Maravalhas JB, Feitosa RM, Pacheco R, Neves KC, Andersen AN (2017) Neotropical savanna ants show a reversed latitudinal gradient of species richness, with climatic drivers reflecting the forest origin of the fauna. Journal of Biogeography 45:248-258. doi: 10.1111/jbi.13113

Wenger SJ, Olden JD (2012) Assessing transferability of ecological models: an underappreciated aspect of statistical validation. Methods in Ecology and Evolution, 3, 260–267.

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation, 122, 99– 112.

Wilson TL, Odei JB, Hooten MB, Edwards TC (2010) Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance. Journal of Applied Ecology, 47, 401-409.

Wisz MS, Hijmans RJ, Li J, Peterson AT, Graham CH, Guisan A (2008) Effects of sample size on the performance of species distribution models. Diversity and Distributions, 14, 763–773.

Wisz MS, Guisan A (2009) Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. BMC Ecology 9:8. doi: 10.1186/ 1472-6785-9-8

Yackulic CB, Chandler R, Zipkin EF, Royle JA, Nichols JD, Grant EHC, Veran S (2013) Presence-only modelling using MAXENT: when can we trust the inferences? Methods in Ecology and Evolution, 4, 236–243.
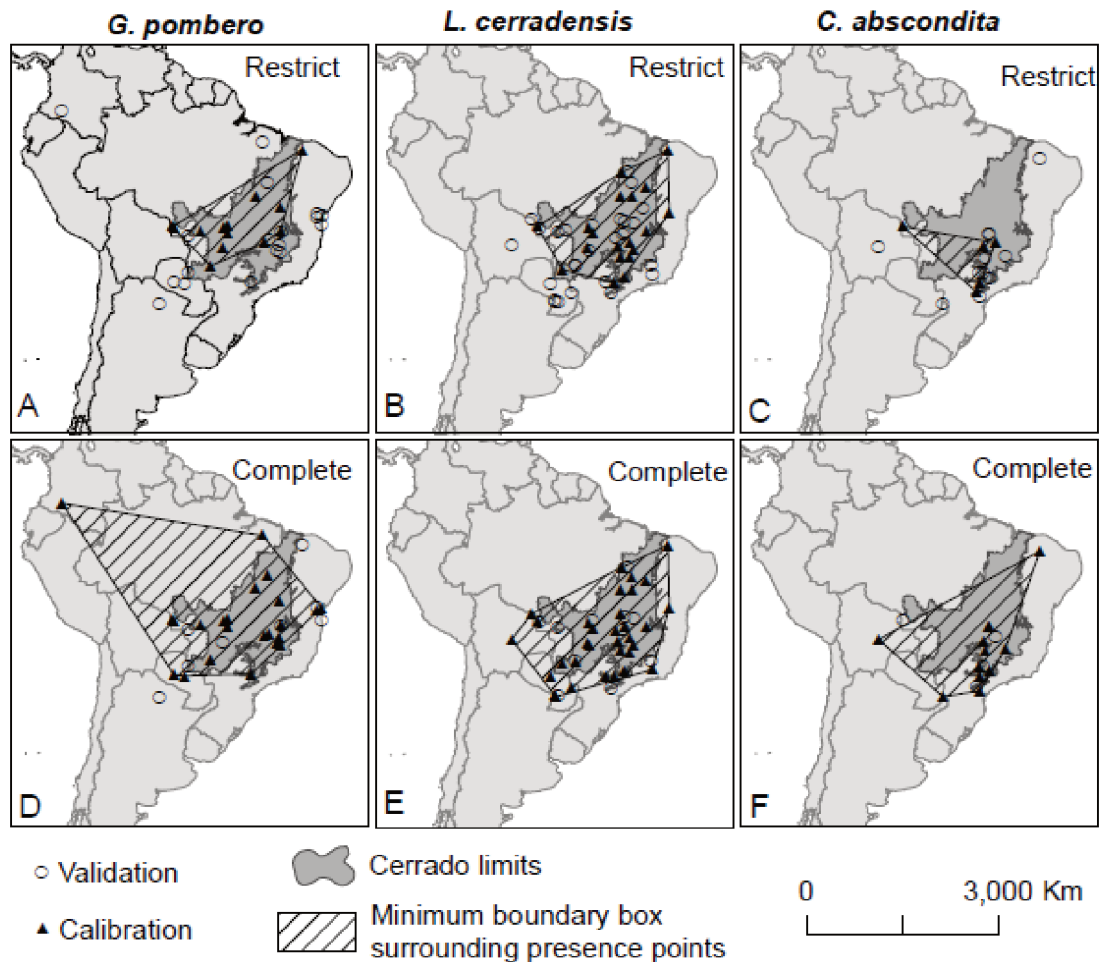
Figures

Figure 1:

Map of South America and limits of the *Cerrado* (savanna) showing the points of occurrence of three savanna/grassland-specialist ant species.

The restrict dataset represents data collected by Vasconcelos et al (2017) in 29 sampling points scattered across *Cerrado*. The complete dataset represents data collected by Vasconcelos et al (2017) and data collected by various authors and available at antmaps.org. The minimum boundary polygon surrounding calibration points represents the area from which pseudo-absences were taken.
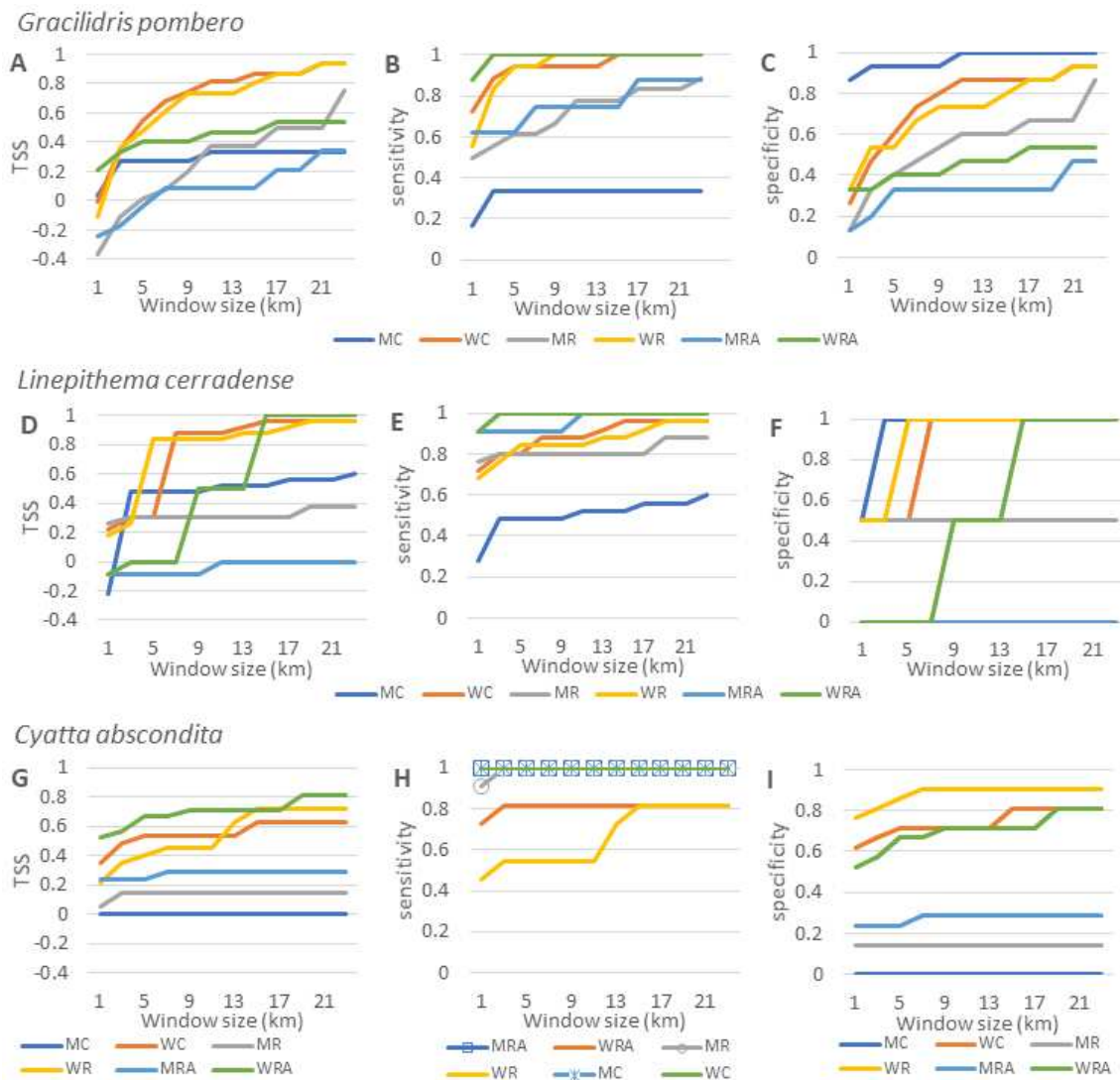
Figure 2:

Evaluation statistics in relation to window sizes.

Shown is the TSS, sensitivity (percentage of correct classification of validation presences), and specificity (percentage of correct classification of absences) statistics. Legend codes refers to predictions of the following methods and datasets: MC: Maxent/Complete dataset; WC: Weights of Evidence/Complete dataset; MR: Maxent/Restrict dataset; WR: Weights of Evidence/Restrict dataset; MRA: Maxent/Restrict dataset including Abundance values; WRA: Weights of Evidence/Restrict dataset including Abundance values.

Figure 3:
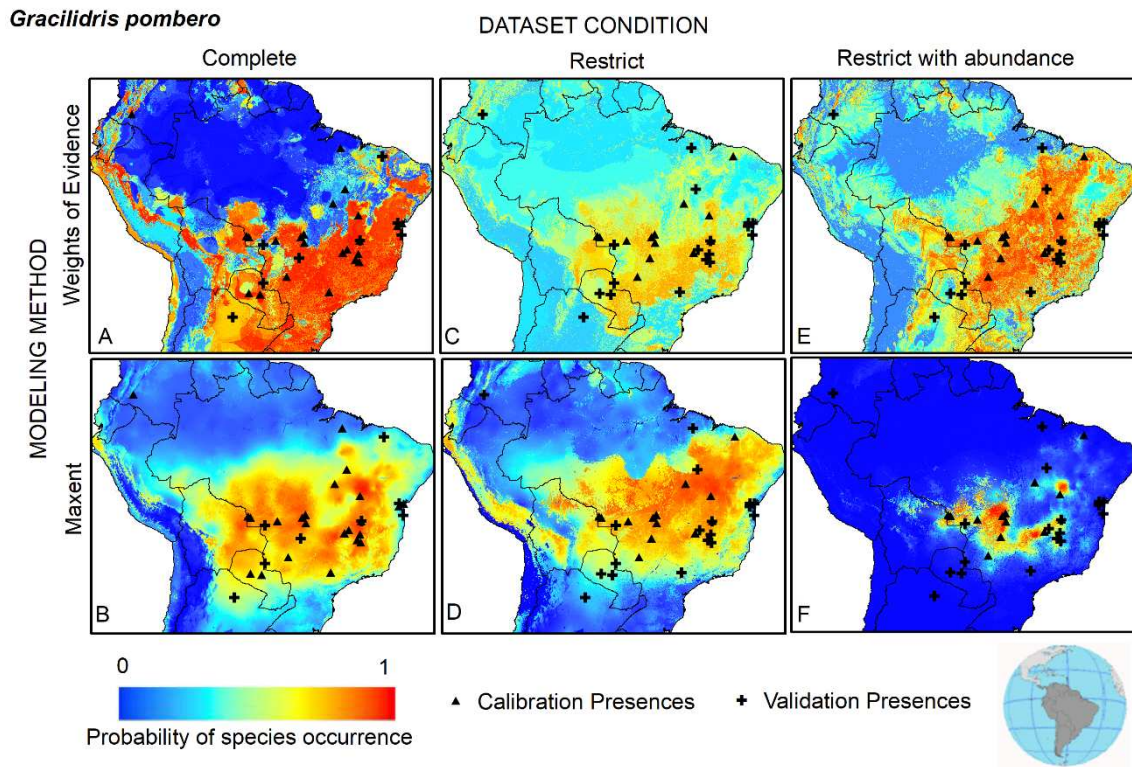
Suitability surfaces for *Gracilidris pombero*.

Panels represent surfaces obtained by applying the following modeling methods to each dataset condition: A) Weights of Evidence (WE), Complete Dataset; B) Maxent, Complete dataset; C) WE, Restrict Dataset; D) Maxent, Restrict Dataset; E) WE, Restrict Dataset including Abundance values; F) Maxent, Restrict Dataset including Abundance values.
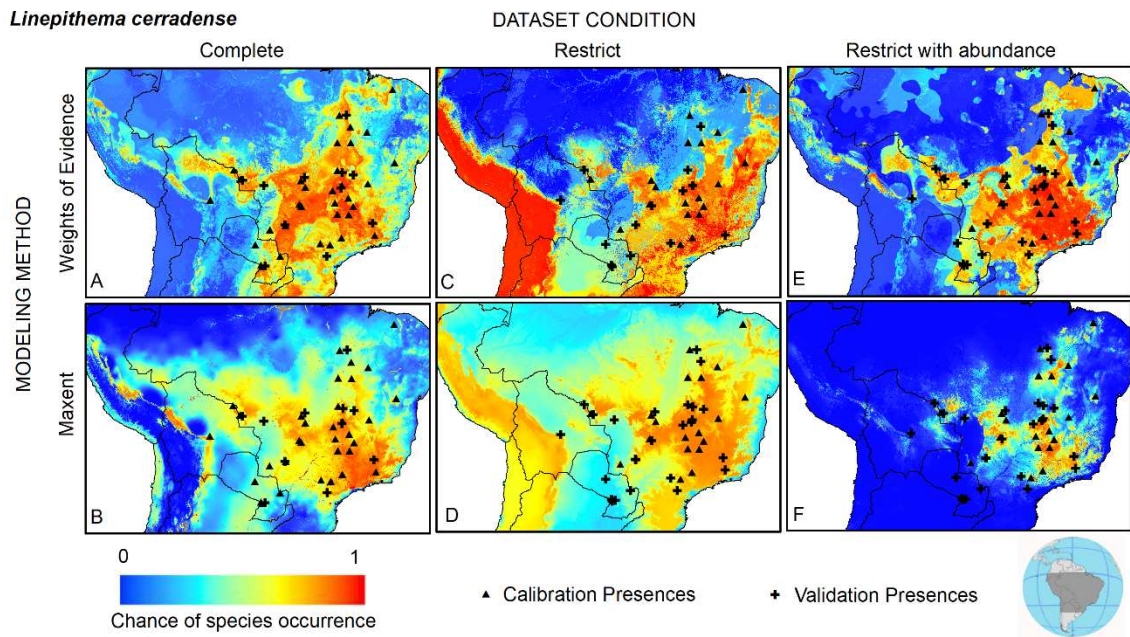
Figure 4:

Suitability surfaces for *Linepithema cerradense*.

Panels represent surfaces obtained by applying the following modeling methods into each dataset condition: A) Weights of Evidence (WE), Complete Dataset; B) Maxent, Complete dataset; C) WE, Restrict Dataset; D) Maxent, Restrict Dataset; E) WE, Restrict Dataset including Abundance values; F) Maxent, Restrict Dataset including Abundance values.

Figure 5:

Suitability surfaces for *Cyatta abscondita*.

Panels represent surfaces obtained by applying the following modeling methods into each dataset condition: A) Weights of Evidence (WE), Complete Dataset; B) Maxent, Complete dataset; C) WE, Restrict Dataset; D) Maxent, Restrict Dataset; E) WE, Restrict Dataset including Abundance values; F) Maxent, Restrict Dataset including Abundance values.
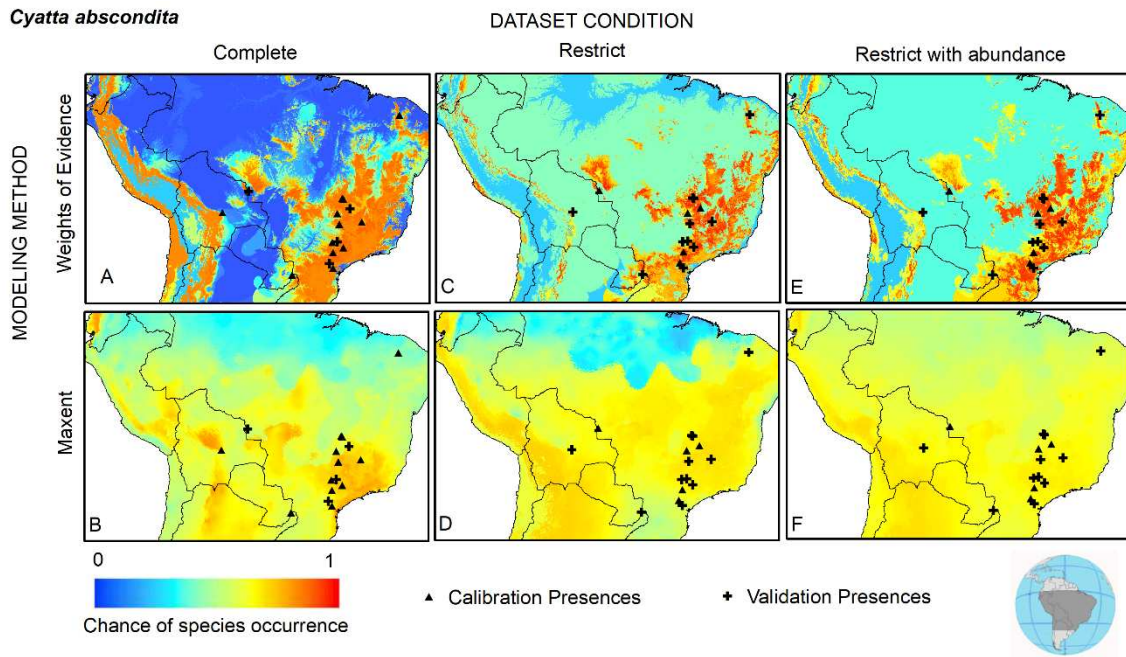
Modelling the effects of climate change on the distribution of ants from South America savannas and grasslands

Abstract

Climate change along with degradation of the natural environment have caused species losses worldwide and, in many cases, changes in ecosystem functioning. It is well known that ectotherm taxa from the tropics are more physiologic vulnerable to climate warming than organisms at higher latitudes. As such, tropical ants, especially those with a restricted geographic distribution may be under high risk. In this context, the main goals of this work were: (1) to estimate current and future (2070) suitability maps for 12 ant species, that occurs exclusively in the savannas or grasslands of South America; (2) to evaluate how effective protected areas are to safeguard these 12 species; (3) to estimate vulnerability degrees of these species based on future suitability surfaces and on protection effectiveness. Current and future suitability surfaces were estimated by means of the Weights of Evidence method based on variables related to climate, soil, elevation and gross primary productivity (GPP). Validation of each species model was based on partial Area under Curve (partial AUC) and sensitivity values (correct classification of presences), estimated for a subset of 25% of presence records not used during the calibration process. Results showed that considering the unlimited dispersal assumption – i.e., ant species will be able to reach novel potential regions by 2070 regardless their distance from current distributions – two species are expected to expand their suitable ranges, reaching around 128%, on average, of their current suitable areas. The other 10 species show 2070 suitable extensions varying from 40% to 91% of areas predicted as suitable nowadays. Considering an opposite situation, in which species will not be able to disperse and, hence their future distribution will be restricted to remaining current areas by 2070, percentual loss of suitable region might be much higher. Given this hypothesis, the extension of suitable ranges may decrease between 13% to 63%. Overall, our study suggests that *Linepithema cerradense* is the least threatened species, due to its 130% increase in suitable area by 2070 and its high representativeness in protected areas, corresponding to 23% in the future. The most vulnerable ones are *Pheidole cyrtostela* and *Blepharidatta conops*, which are expected to maintain between 13% and 16% of current suitable regions. Since these are species that, until today, have been recorded exclusively in some vegetation types of South America, their extinction may mean a world biodiversity loss.

Introduction

Climate change worldwide has been the main cause of observed modifications on distribution ranges of many species (Urban, 2015; Pounds et al., 2006), which, in turn, alters community structure and triggers functional changes in species assemblages and associated ecosystem processes (Pecl et al., 2017; Tiede et al, 2017). In spite of a relatively greater increase in temperature to be expected in higher latitudes rather than in lower ones (IPCC, 2014), for ectotherm taxa, the effect of climate warming might be worst in the tropics, since ectotherms inhabiting low latitudes are more physiologically vulnerable to climate warming than organisms at higher latitudes (Deutsch et al., 2008; Sunday et al., 2011). Among these taxa, ants (Hymenoptera: Formicidae) constitute a relevant functional group. They are important for food web interactions and are involved in numerous ecosystem processes (Del Toro et al., 2015), such as predation, nutrient cycling, soil turnover and seed dispersal (Underwood and Fisher, 2006; Parr et al, 2016). Potential changes in ant species distributions caused by climate warming may significantly alter ant-mediated ecosystem processes and services (Del Toro et al., 2015).

Ant species richness and composition is often correlated with climatic gradients (del Toro et al., 2015; del Toro et al, 2019; Vasconcelos et al, 2017) and vegetation structure (Hoffmann, 2010; Underwood and Fisher, 2006; Gibb and Parr, 2013; Gibb et al., 2015; Wiescher et al., 2012, Tiede et al., 2017). Diamond et al (2012) showed that climate warming may have a negative effect on tropical ant species due to their low thermal tolerance. Climate effects may act even more rapidly on endemic and range-restricted ant species, not only because they might present even smaller temperature buffers but also due to their lower ability to track environmental changes (Parmesan, 2006; Domisch et al., 2013). In the tropics, the additional effect of the ongoing deforestation (Strassburg et al., 2017) may increase ant vulnerability even more. The South America tropical savanna - that composes the second largest Biome in South America, the *Cerrado* - have lost 46% of its natural cover and only 19% remain undisturbed (Strassburg et al., 2017). In this sense, anticipating the impacts that future changes may have on tropical ants constitutes an important issue to help developing conservation strategies.

Species distribution models (SDMs) are useful tools to estimate future species distributions given environmental and climate change and, hence, constitutes a key procedure in conservation planning (Wilson et al., 2005). SDMs basically identifies

relationships between environmental variables and species occurrence in order to predict species distribution ranges. Due to the scarce availability data (Jenkins et al., 2013), only recently SDMs have been applied to predict future consequences of climate change on of insects (Nóbrega & De Marco 2011; Giannini et al., 2013; Silva et al., 2013; Martins et al., 2014; De Marco et al, 2015; Pires et al., 2018; Silva et al., 2019). Regarding ants, there are studies evaluating the possible effects of future climate on species richness and composition (Del Toro et al., 2015; Tiede et al., 2017; Diamond et al., 2012), but only one have dedicated to estimate their future spatial distributions (del Toro et al, 2015).

An extensive survey of the Neotropical savanna ants (Vasconcelos et al., 2017) revealed that there are at least 12 range-restricted species that occurs exclusively, or predominantly, in the *Cerrado* biome (Rodrigo Feitosa, personal communication). In this context, the main goals of this work were: (1) to assess the effect of climate change on the distribution of the 12 range-restricted savanna ant species; (2) to evaluate how effective protected areas are to safeguard these 12 species; (3) to estimate vulnerability degrees of these species based on future suitability surfaces and on protection effectiveness.

Material and Methods
Data

The 12 ant species analyzed here represent a subset of the 47 species listed as grassland/savanna specialists by Vasconcelos et al (2017). The selected species were chosen because their known distribution is largely coincident with the distribution of the *Cerrado* biome. In addition, these species have a relatively well-resolved taxonomy, and thus are less prone to identification errors (Rodrigo M. Feitosa, personal communication).

Among the selected ant species, nine belong to the Attini tribe (subfamily Myrmicinae). These are *Cyatta abscondita* Soza-Calvo et al, 2013; *Xerolitor explicatus* Kempf 1968; *Mycetagroicus cerradensis* Brandão & Mayhé-Nunes 2001; *Trachymyrmex dichrous* Kempf 1967; *Blepharidatta conops* Kempf 1967; *Sericomyrmex maravalhas* Ješovnik & Schultz, 2017; *Sericomyrmex scrobifer* Forel 1911, *Cephalotes betoi* de Andrade & Urbani, 1999; and *Pheidole cyrtostela* Wilson 2003. The three remaining species are *Linepithema cerradense* Wild 2007, *Linepithema aztecoides* Wild 2007, both in the Dolichoderinae subfamily, and *Ectatomma planidens* Borgmeier 1939, a representative of the subfamily Ectatominae.

Occurrence numbers for these species varied from six to 44 records (table 1). The data (composed of species presences records) were split into calibration (or training) and validation (or test) data. Validation dataset corresponded to 25% of total presences and were obtained through a systematic sampling ordered by latitude values of presences. The remaining 75% composed the calibration group.

The variables used to model each species suitability across the space took into account that the occurrence of different ant species is, to a large extent, determined by variation in climate, elevation gradients, vegetation cover and soils (Del Toro et al., 2015, 2018; Vasconcelos et al., 2017). Climatic data was obtained from WorldClim version 1.4 (Hijmans, 2005) in a spatial resolution of, approximately, 0.041 degrees (~4.5 km). Climatic conditions constitute of 19 maps representing combination of means, extremes, and seasonality in rainfall and temperature. Future climate corresponded to projections of these same 19 variables also estimated by WorldClim version 1.4 and reflects projected climate by 2070 in the worst IPCC scenario (RCP 8.5, IPCC 2014). Altitude were derived from SRTM (Shuttle Radar Topography Mission) maps. To account for vegetation characteristics and also the energy available in the system, I used Gross Primary Productivity (GPP) maps of present-day and 2070 developed by Madani et al. (2018). Finally, to account for soil characteristics that might interfere in building nests, I have downloaded soil data from soilgrids.org. These data show estimated sand and clay content at seven depth levels each, varying from the surface to 2 meters deep. In total, I had 37 candidate maps of potential predictors of ant distributions.

The protected areas network was downloaded from the World Database on Protected Areas (WDPA 2015, http://www.wdpa.org). I have considered here only protected areas whose categories varied from I-IV according to the codes of the International Union for Conservation of Nature (IUCN).

Methods

The analysis applied here constitutes on three steps. In the first one, I follow the method explained in Chapter 1 to fit probability models for occurrence of each one of the 12 ant species selected for this study. After validating current suitability surfaces, fitted models were used to generate future distributions of these ant species. In the last step, I accounted for the expected changes in suitable areas caused by climate change and evaluated the match between predicted surfaces and the protected areas network.

Considering modelling suitability surfaces, Weights of Evidence, hereafter referred as WE, (Agtberg and Bonhan-Carter, 1990; Bonham-Carter, 1994; Soares Filho et al., 2009) was selected because most ant species have few and restricted presence records, two conditions against which WE proved to have superior ability than Maxent (Philips et al., 2006), a technique commonly used to model few data. For each variable, WE calculates a series of weights, which are combined into a single equation that generates the suitability maps (Bonham-Carter, 1994). The weights describe the relationship between the variable and the occurrence of the species, and, hence can be a basic guide to suggest species preferences. Model calibration for each species was conducted following the exact steps described in Chapter 1 taking into account calibration presence records. From the 37 possible explaining variables, final models of each species contained just the most significant and not correlated ones, also according to the approach suggested in Chapter 1.

I evaluated each model using two different evaluation metrics. The first one consisted on partial AUC (area under curve; Peterson and Soberon, 2008), which was run by means of "roc" function of the pROC package in R (R Core Team, 2017). This function was applied in calibration and validation presence points plus an amount of spatially random points, correspondent to ten times the number of total presences. As a second evaluation metric, I evaluated sensitivity values of validation datasets across different scales, varying from 1 cell size (~4.5 km) to 11 cells (~50 km). The analysis of several scales allows to account as correct classification the case of a cell predicted as presence that fall not in the exact position of an observed record but within its neighborhood. A neighborhood analysis like this is relevant mainly in situations in which explaining variables were not directly measured but interpolated and, hence, small space displacements might occur. To estimate sensitivity, it is necessary, firstly, to determine expected presences and absences. The value used to determine if a probability cell constituted on an estimated presence or absence corresponded to the threshold that maximized the sum sensitivity plus specificity by means of the function Find.Optim.Stat of the package Biomod2, in R (R Core Team, 2017). Although the database of Vasconcelos et al (2017) shows sites where the species was not detected, I opted for not using these places as absence points as there is a large chance that they do not represent true absences due to the fact that observed abundance was very low (in many cases equal to one) for the majority of the species.

Also based in this threshold, I calculated suitable and non-suitable areas and, then - given suitability maps for 2070 obtained by applying fitted models to 2070 data - estimated expected habitat loss or gain between nowadays and 2070. As a last step, suitable areas for each species were compared to the protected areas network in order to estimate the vulnerability degree of each species. Classification of species into vulnerability degrees was based on a vulnerability graph, similar to Velaszco (2019). Here, vulnerability graph shows, in the Y axis, the percentage of current suitable areas that are expected to be maintained as suitable in 2070, and in the X axis, the percentual loss of suitable regions inside protected areas.

Results

General Species preferences

The most important variables describing the potential distribution of the studied species, according to Weights of Evidence, were rainfall, temperature and GPP. All fitted models contained rainfall and temperature variables and nine of them showed also GPP as a crucial predictor (table 2).

Considering the eight ant species for which GPP was a significant variable (table 2), their GPP interval with positive weights ranged between 999 – 2261 g m-2year-1 (Fig. 1A), a range that corresponds to a third part of the total range observed in South America. As positive weights reflect an association between the explaining variable and the modeled event, our analysis suggest that the general GPP interval preferred by the ant species analyzed here might be 999 – 2261 g m-2year-1. A closer analysis of GPP weights for each species revealed that each one shows a different tolerance to vegetation conditions, since specific ranges and weights varied significantly between species (Fig. 1A). *Trachymyrmex dichrous* showed the broadest interval of preferred GPP, since it presented the largest spectrum of GPP positive weights (Fig. 1A). On the other hand, *Pheidole cyrtostela* may be the most demanding species concerning GPP values due to its small range of GPP (Fig. 1A). The four species in which GPP was not a significant variable were *Sericomyrmex scrobifer*, *Sericomyrmex maravalhas*, *Mycetagroicus cerradensis* and *Cyatta abscondita*.

The rainfall related variable most used in final fitted models was annual precipitation. It was part of the models of seven species (*Blepharidatta conops*, *Linepithema cerradense*, *Linepithema aztecoides Xerolitor explicatus*, *Sericomyrmex*

scrobifer, *Ectatomma planidens* and *Trachymyrmex dichrous*, table 2), whose preferred values ranged between the 116 – 2015 mm interval (Fig. 1B). By means of comparison, precipitation range of 99% of South America territory correspond to 0 – 4000 mm. The species most adapted to dry conditions was *Xerolitor explicatus*, whose positive precipitation weights begun at 116 mm. At the same time, this species showed the broadest range, from 116 to 1359 mm (Fig. 1B). Among the five species whose final models did not include annual precipitation, four included other rainfall related variables. *Pheidole cyrtostela*, *Sericomyrmex maravalhas* and *Mycetagroicus cerrandensis* were correlated to precipitation of wettest month, showing positive weights from 173 to 296 mm, 206 to 362 mm and 178 to 280 mm, respectively. The final model of *Cephalotes betoi* took into account the precipitation of driest quarter, with positive weights ranging from 24 to 64 mm.

General temperature preferences could not be seen in this study, since each species model showed a different arrangement of significant temperature variables. In contrast to rainfall data, which had only four variables included in the fitted models, temperature characteristics of final models included nine temperature variables (Table 2). Five of these variables were exclusive of just one of the fitted models, three variables were present in two models and just one, the maximum temperature of warmest month was included in three models.

Surprisingly, just one species, *Trachymyrmex dichrous*, had soil characteristics in the final model (table 2). Elevation took part of the models of five species, listed along with their preferred altitude ranges: *Cyatta abscondita*, 360 to 1250 m, *Ectatomma planidens*, 450-1450 m, *Sericomyrmex scrobifer*, 608-1353 m, *Mycetagroicus cerradensis*, 607-1192 m, and *Trachymyrmex dichrous*, 647-1362 m.

<u>Fitted suitability surfaces</u>

Our data suggest that, under the assumption of unlimited dispersal (Merow et al. 2013; Owens et al. 2013), the total extension of suitable areas might decrease for 10 species by 2070 (Figs. 2 and 3). This assumption consider that ant species will be able to reach novel potential regions by 2070 regardless their distance from current distributions. The two species that might expand their suitable areas are *Linepithema cerradense* and *Mycetagroicus cerradensis* (Fig 3), which are expected to have 131% and 125% of their current areas by 2070, respectively. Their expansion areas tend to shift towards the north

and south of their current distributions (Figs 2B and 2K). The huge expansion of suitable conditions for *Linepithema cerradense* towards the Amazon (Fig 2B) was caused by the joint effect of drier conditions and diminished productivity of vegetation in this region by 2070. Weights of evidence analysis revealed that current preferred ranges of GPP for *Linepithema cerradense* varies between 1243 and 2055 g m-2year-1 (Fig. 1A) and also that this species currently occupies regions with annual precipitation varying from 1328 to 1836 mm (Fig 1B). By 2070, these ranges are expected to occur in the southeast Amazon. The other species may have suitable areas corresponding to 40% - 91% of the total area predicted as currently suitable (Fig. 3). The species with smaller percentual loss in areal size, varying from 9% to 20%, were *Xerolitor explicatus*, *Cephalotes betoi* and *Trachymyrmex dichrous*. *Ectatomma planidens*, *Blepharidatta conops* and *Sericomyrmex scrobifer* may have future suitable areas varying between 65-67% of suitable areas observed today. The highest percentual losses are expected to *Sericomyrmex maravalhas*, *Linepithema aztecoides*, *Cyatta abscondita* and *Pheidole cyrtostela* , which may show 2070 suitable ranges between 45%- 60% smaller than current suitable areas.

Evaluating only the percentage of current suitable areas expected to be maintained by 2070 - and, hence, simulating an opposite situation of the unlimited dispersal assumption – all species are expected to lose suitable areas. *Linepithema cerradense* and *Mycetagroicus cerradensis*, two species that showed increase in suitable ranges under unlimited dispersal assumption, are expected to maintain 61% and 38% of their current suitable areas by 2070. Among the 10 species expected to have total suitable areas in 2070 smaller than the current ones (Fig. 3), the species of ants with the highest maintenance rates of suitable regions are *Sericomyrmex scrobifer*, *Ectatomma planidens* and *Trachymyrmex dichrous* (Figs. 2J, 2L and 2D). These species are expected to maintain, respectively, 45%, 50% and 63% of their current suitability areas by 2070. *Blepharidatta conops* and *Pheidole cyrtostela* presented the lowest expected proportion of current suitable areas to be conserved by 2070 (Figs. 2A and 2G), corresponding to 16% and 13%, respectively. The species with intermediate rates of maintenance of current suitable areas were *Linepithema aztecoides*, *Cyatta abscondita*, *Xerolitor explicatus* and *Cephalotes betoi* (Figs. 2C, 2E, 2F and 2H), with preservation percentages ranging from 17% to 32%.

Considering that occupation of entire new suitable areas by 2070 (unlimited dispersion) might not be feasible for the majority of the species, and also that total limited

dispersion is not likely to occur, I have analyzed an intermediate situation. I have considered that effective new potential areas - the ones that might be reached by the species - will be formed by contiguous regions spanning the 2070 suitable area that is connected (share borders) to remaining parts of current suitable areas by 2070. Within this context, *Linepithema cerradense* might be the species with the greatest chance to successfully persist in South America, since its new potential area predicted by 2070 shares a great border with areas expected to be suitable both in present-day and in the future (Fig. 2B). On the opposite, new future areas that shares borders only with present-day suitable regions might not be effective to shelter the species. This is because we do not know when proper climate conditions for each species will be disrupted and, more importantly, until when species are able to persist in a determined area under a changing climate. For example, if proper conditions for a given species extinguishes by 2040, then the existence, in the future, of a suitable region next to the present-day distribution might be not useful. In this context, once more, our study suggests that *Blepharidatta conops* and *Pheidole cyrtostela* are expected to be the most impaired species because, besides showing the smallest overlap between current and future suitability areas, a significant part of 2070 new potential suitable areas are located far beyond the former cited overlapping area (Fig. 2A and 2G). Another species that might be seriously impacted by future changes is *Xerolitor explicatus*. This species showed the most restricted distribution range for both present-day and future, in addition, its overlapping area between present-day and 2070 is not connected to new suitable regions (Fig. 2F).

Still regarding connection between future and present-day distributions, the species *Ectatomma planidens*, *Trachymyrmex dichrous*, *Sericomyrmex maravalhas*, *Sericomyrmex scrobifer*, *Linepithema aztecoides*, *Cyatta abscondita*, *Cephalotes betoi* and *Mycetagroicus cerradensis* might, in the future, stay restricted to the reduced amount of suitable regions constituted by the overlapping present-day and 2070 areas, since new 2070 suitable regions for these species are really small or located far beyond current areas (Fig 2L, 2D, 2I, 2J, 2C, 2E, 2H, 2K).

<u>Evaluating fitted models</u>

Evaluation of fitted probability maps, the ones used to produce suitability regions presented in Fig 2, showed satisfying results for all species. Partial AUC values varied between 0.83 and 0.95 (Fig. 4A) and correct prediction of presences (sensitivity) ranged from 67% to 100% in a cell by cell comparison (Fig 4B). If we increase evaluating

window size to 5 x 5 cells, corresponding to approximately 22.5 km x 22.5 km, the final model for *Xerolitor explicatus* classifies correctly 96% of the presences. If we consider a bigger window, of 11 x 11 cells – i.e. 50 km - the *Xerolitor explicatus* model predicts all presences and the model for *Sericomyrmex scrobifer* 83% of them. By this scale, the only models that did not reach 100% of accuracy were the ones of *Pheidole cyrtostela*, *Linepithema aztecoides* and *Trachymyrmex dichrous*. Models for *Pheidole cyrtostela* and *Linepithema aztecoides* reached maximum accuracy of presences in window sizes of 13 x 13 cells and in a 15 x 15 cells, respectively. The only model that did not reach 100% of correct classification of presences within a 15 x 15km window was the one for *Trachymyrmex dichrous*.

Comparing suitable regions with protected areas network

All species was represented in at least one protected area. The majority of the species have between 10% and 15% of their current suitable distribution overlaid to protected areas (Fig. 5). If we consider predicted suitable area by 2070, under the assumption of unlimited dispersion, this percentage is expected to increase for seven species, the ones whose regions with proper conditions are predicted to go towards the north, a region that contains a much larger area under protection. Nevertheless, a more conservative assessment as suggested by opposite conditions of dispersal - considering that species will not be able to migrate, and, hence, taking into account only current suitable areas that will persist with proper conditions in the future - revealed that percentage of protection in the future might be, on average, as low as 5%.

Level of vulnerability for each species can be classified into four categories, varying from I, the less vulnerable group, to IV, the most threatened one (Figure 6). Group I comprised two species, *Linepithema cerradense* and *Trachymyrmex dichrous,* which are expected to maintain more than 60% of current suitable areas by 2070 and have the lowest percentual loss inside protected areas (figure 6). Group II comprises the species *Ectatomma planidens*, *Sericomyrmex scrobifer* and *Mycetagroicus cerradensis* which are predicted to maintain between 39% and 50% of current estimated distributions by 2070 and are expected to loss between 63% and 65% inside protected areas by this same period (Figure 6). Group III includes *Cyatta abscondita*, *Sericomyrmex maravalhas*, *Xerolitor explicatus, Cephalotes betoi* and *Linepithema aztecoides,* whose maintenance percentages varied from 25% to 32% and percentual loss inside protected areas might reach 69% to 85%. Group IV consists of the most threatened ones: *Pheidole cyrtostela*

and *Blepharidatta conops*, which are expected to maintain only between 13% and 16% of their current estimated distributions and are estimated to loose between 79% and 88% of suitable regions inside protected areas (Figure 6).

Discussion

This is the first study to evaluate the effects of future climate change on the distribution of South American ant species. Level of vulnerability for each species was classified into four categories, varying from I, the less vulnerable group, to IV, the most threatened one. Within all species, *Linepithema cerradense* was the least vulnerable one, since our analysis indicated an increase of 130% in the suitable area by 2070, of which 23% is expected to be located in protected areas. In addition, this species presented a wide contiguous zone comprising 2070 suitability together with remaining current suitable areas. The existence of a contiguous current-2070 area may facilitate the migration process along the years. In contrast, for all the other species, a contiguous current-2070 area is small or inexistent (Fig. 2).

Our study suggests that species from groups II to IV might have a small amount of future range represented in protected areas, corresponding to an average of 10%, a value that could be diminished to a mean of 5% if we consider the limited dispersal assumption. This conclusion is in accordance to other studies regarding future distributions of insects in South America, which have also predicted a very small coverage of protected areas (Martins et al., 2015; Pires et al., 2018; Silva et al., 2019). Another factor that must be taken into account is that savannas and grasslands in South America contain a very small portion under protection zones (Strassberg et al., 2017; Bilenca and Miñarro 2004). For example, grasslands areas under legal protection in Argentina and Southern Brazil ranges between 1 to 1.5% (Bilenca and Miñarro 2004) and, in *Cerrado*, just 2.89% fall into the category of fully protected land (Strassberg et al., 2017).

In general, fitted models included explaining variables that have already been identified as correlated with the distribution of ants. All fitted models presented at least one temperature variable and six of them showed between 2 and 5 temperature variables, indicating that model fitting process could capture the high sensitivity of tropical ants to temperature conditions (Diamond et al., 2012). The relationship between ant species

richness and precipitation has already been pointed out (Del Toro et al, 2018), and, here, rainfall has shown to be strongly correlated to species presences, since 11 models presented at least one variable related to precipitation. The joint effect of temperature, rainfall and GPP guided species suitable conditions towards the north of Brazil, more specifically, to the south of Amazon. However, although GPP values in this expanded range are likely to be similar to the ones currently observed in *Cerrado*, we do not know if vegetation structure by that time will allow the presence of species.

Here, I have dedicated to estimate climatically suitable areas for individual ant species. These surfaces should be seen as basic indicators for species maintenance in future years, since their effective survival depends on other factors which were not able to be measured here. A different climate might alter species in different ways, depending on their functional traits, diet and foraging strategy (Gibb and Parr, 2013; Gibb et al., 2015; Wiescher et al., 2012; Tiede et al., 2017). As a consequence, population and community patterns are likely to change and the new emerging interactions may cause substantial changes in species distributions, which were not able to be accounted in this study. For example, since climate change may impact fungus (Rodder et al., 2010; Shrestha & Bawa 2014), six species analyzed here are expected to suffer the additional effect that future climate may have on their symbiotic fungus.

In this context, future studies should aim not only to model species distributions, but also to explore how the complex relationships within a community might be altered as a joint effect of climate and land use changes. Considering there is a lack of systematic nation-wide surveys that could be used as basis to such studies, especially in developing countries (El Gabbas & Dormann, 2017), the first step to model communities is to create systematic species databases within the community level. Given this database, it would be able to develop monitoring programs that, beyond alerting us to biodiversity losses, would suggest functional changes in species assemblages and associated ecosystem functioning and services.

Conclusion

Overall, our study suggests that *Linepithema cerradense* is the least threatened species, due to its 130% increase in suitable area by 2070 and its high representativeness in protected areas, corresponding to 23% in the future. The most vulnerable ones are *Pheidole cyrtostela* and *Blepharidatta conops*, which are expected to maintain between

13% and 16% of current suitable regions. Since these are species that, until today, have been recorded exclusively in some vegetation types of South America, their extinction may mean a world biodiversity loss.

References

Agterberg FP, Bonham-Carter GF (1990) Deriving weights of evidence from geoscience contour maps for the prediction of discrete events. Pages 381–395 in XXII International Symposium. Volume 2. APCOM, Berlin, Germany.

Bilenca D, Miñarro F (2004) Identificación de Áreas Valiosas de Pastizal (AVPs) en las Pampas y Campos de Argentina, Uruguay y sur de Brasil. Fundación Vida Silvestre, Buenos Aires.

Bonham-Carter G (1994) Geographic information systems for geoscientists: modelling with GIS. Pergamon, Oxford, UK.

Borgmeier T (1939) Nova contribuição para o conhecimento das formigas neotropicais (Hym. Formicidae). Rev. Entomol. (Rio J.) 10: 403-428.

Brandão CRF, Mayhé-Nunes AJ (2001). A new fungus-growing ant genus, Mycetagroicus gen. n., with the description of three new species and comments on the monophyly of the Attini (Hymenoptera: Formicidae). Sociobiology 38(3B): 639-665 (page 644, figs. 1-7 worker described)

de Andrade ML, Baroni Urbani C (1999) Diversity and adaptation in the ant genus Cephalotes, past and present. Stuttg. Beitr. Naturk. Ser. B (Geol. Paläontol.) 271: 1-889 (page 347, figs. 153, 154, 386 soldier, worker described).

De Marco Jr. P, NObrega CC, Souza RR et al (2015) Modeling the distribution of a rare Amazonian odonate in relation to future deforestation. Freshwater Science, 34(3): 1123-1132.

Del Toro I, Ribbons RR, Hayward J et al. (2019) Are stacked species distribution models accurate at predicting multiple levels of diversity along a rainfall gradient? Austral Ecology, 44(1): 105-113.

Del Toro I, Silva RR, Ellinson AM (2015) Predicted impacts of climatic change on ant functional diversity and distributions in eastern North American forests. Diversity and Distributions, 21: 781–791.

Deutsch CA, Tewksbury JJ, Huey RB, et al. (2008) Impacts of climate warming on terrestrial ectotherms across latitude. Proceedings of the National Academy of Sciences of the United States of America, 105: 6668–6672.

Diamond SE, Sorger DM, HULCR J (2012) Who likes it hot? A global analysis of the climatic, ecological, and evolutionary determinants of warming tolerance in ants. Global Change Biology 18(2): 448-456.

Domisch S, Araújo MB, Bonada N, et al (2013) Modelling distribution in European stream macroinvertebrates under future climates. Global Change Biology, 19:752–762.

El-Gabbas A, Dormann CF (2017) Improved species-occurrence predictions in data-poor regions: using large-scale data and bias correction with down-weighted Poisson regression and Maxent. Ecography, https://doi.org/10.1111/ecog.03149

Forel A (1911). Ameisen des Herrn Prof. v. Ihering aus Brasilien (Sao Paulo usw.) nebst einigen anderen aus Südamerika und Afrika (Hym.). Dtsch. Entomol. Z. 1911: 285-312 (page 296, worker described)

Giannini TC, Acosta AL, Silva CI et al. (2013) Identifying the areas to preserve passion fruit pollination service in Brazilian Tropical Savannas under climate change. Agriculture, Ecosystems and Environment, 171: 39–46.

Gibb H, Parr CL (2013) Does structural complexity determine the morphology of assemblages? An experimental test on three continents. PLoS One 8(5): e64005.

Gibb H, Stoklosa J, Warton DI et al. (2015) Does morphology predict trophic position and habitat use of ant species and assemblages? Oecologia, 177: 519–531.

Hijmans RJ, Cameron SE, Parra JL et al. (2005) Very high resolution interpolated climate surfaces for global land areas. International journal of climatology, 25(15): 1965-1978.

Hoffmann BD (2010) Using ants for rangeland monitoring: global patterns in the responses of ant communities to grazing. Ecological Indicators 10: 105–111.

IPCC (2014). Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: Intergovernmental Panel on Climate Change.

Jenkins CN, Pimm SL, Joppa Lucas N (2013) Global patterns of terrestrial vertebrate diversity and conservation. Proceedings of the National Academy of Sciences of USA, 110: 2602-2610.

Ješovnik A, Schultz TR (2017) Revision of the fungus-farming ant genus Sericomyrmex Mayr (Hymenoptera, Formicidae, Myrmicinae). ZooKeys. 670:1–109.

Kempf WW (1967) New ants from southeastern and central Brazil (Hymenoptera, Formicidae). Stud. Entomol. 9: 121-128 (page 126, figs. 7-9 worker described)

Kempf WW (1968) Miscellaneous studies on Neotropical ants. IV. (Hymenoptera, Formicidae). Stud. Entomol. 11: 369-415 (page 401, figs. 14-16 worker described)

Madani N, Kimball JS, Ballantyne AP et al. (2018) Future global productivity will be affected by plant trait response to climate. Science Reports 8: e2870.

Martins AC, Silva DP, de Marco Jr P (2014) Species conservation under future climate change: the case of Bombus bellicosus, a potentially threatened South American bumblebee species. Journal of Insect Conservation, 19(1): 33-43.

Nóbrega CC, De Marco Jr. P (2011) Unprotecting the rare species: a niche-based gap analysis for odonates in a core *Cerrado* area. Diversity and Distributions, 17: 491–505

Parmesan C (2006) Ecological and evolutionary responses to recent climate change. Annual Review of Ecology Evolution and Systematics, 37: 637–669.

Parr CL, Eggleton P, Davies AB, et al. (2016). Suppression of savanna ants alters invertebrate composition and influences key ecosystemprocesses. Ecology 97: 1611–1617.

Pecl, GT, Araújo MB, Bell JD, et al. (2017) Biodiversity redistribution under climate change: Impacts on ecosystems and human wellbeing. Science, 355: 9214.

Peterson AT, Soberon MPJ (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modeling. Ecological Modelling, 213: 63–72.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. Ecological Modelling 190: 231-259.

Pires MM, Périco E, Renner S et al. (2018) Predicting the effects of future climate change on the distribution of an endemic damselfly (Odonata, Coenagrionidae) in subtropical South American grasslands. Journal of Insect Conservation, 22: 303–319.

Pounds JA, Bustamante MR, Coloma LA et al (2006) Widespread amphibian extinctions from epidemic deseases driven by global warming. Nature 439: 161-167.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rodder D, Kielgast J, Lotters S (2010) Future potential distribution of the emerging amphibian chytrid fungus under anthropogenic climate change. Diseases of Aquatic Organisms, 92: 201–207.

Shrestha UB, Bawa KS (2014) Impact of climate change on potential distribution of Chinese caterpillar fungus (Ophiocordyceps sinensis) in Nepal Himalaya. PLoS ONE 9: e106405

Silva D, Aguiar A, Melo G et al. (2013) Amazonian species within the *Cerrado* savanna: new records and potential distribution for Aglae caerulea (Apidae: Euglossini). Apidologie, Springer Verlag, 44 (6): 673-683.

Silva DP, Dias AC, Lecci LS et al (2019) Potential Effects of Future Climate Changes on Brazilian Cool-Adapted Stoneflies (Insecta: Plecoptera). Neotropical Entomology, 48: 57-70.

Soares-Filho BS, Rodrigues HO, Costa WL (2009). Modeling Environmental Dynamics with Dinamica EGO. https://www.csr.ufmg.br/dinamica/dokuwiki/doku.php?id=tutorial:start. (Acessed May, 2018).

Sosa-Calvo J, Schultz TR, Brandão CRF et al (2013). *Cyatta abscondita*: Taxonomy, Evolution, and Natural History of a New Fungus-Farming Ant Genus from Brazil. PLoS ONE 8(11): e80498.

Strassburg BBN, Brooks T, Feltran-Barbieri R, et al. (2017) Moment of truth for the *Cerrado* hotspot. Nature Ecology and Evolution, 1: 1–3.

Sunday JM, Bates AE, Dulvy NK (2011) Global analysis of thermal tolerance and latitude in ectotherms. Proceedings of the Royal Society of London Series B Biological Sciences, 278, 1823–1830.

Tiede Y, chlautmann J, Donoso DA (2017) Ants as indicators of environmental change and ecosystem processes. Ecological Indicators, 83: 527–537.

Underwood EC, Fisher BL (2006) The role of ants in conservation monitoring: if, when, and how. Biological Conservation, 132:166–182.

Urban MA (2015) Accelerating extinction risk from climate change. Science, 348 (6234): 571-573.

Vasconcelos HL, Maravalhas JB, Feitosa RM (2017) Neotropical savanna ants show a reversed latitudinal gradient of species richness, with climatic drivers reflecting the forest origin of the fauna. Journal of Biogeography, 45: 248–258.

Wiescher PT, Pearce-Duvet JMC, Feener DH (2012) Assembling an ant community: species functional traits reflect environmental filtering. Oecologia, 169: 1063–1074.

Wild AL (2007) Taxonomic revision of the ant genus Linepithema (Hymenoptera: Formicidae). University of California Publications in Entomology. 126:1-159.

Wilson EO (2003) Pheidole in the New World: A dominant, hyperdiverse ant genus. Harvard University Press, Cambridge, MA. (page 184, fig. major, minor described).

Wilson KA, Westphal MI, Possingham HP, Elith J (2005) Sensitivity of conservation planning to different approaches to using predicted species distribution data. Biological Conservation, 122: 99– 112.
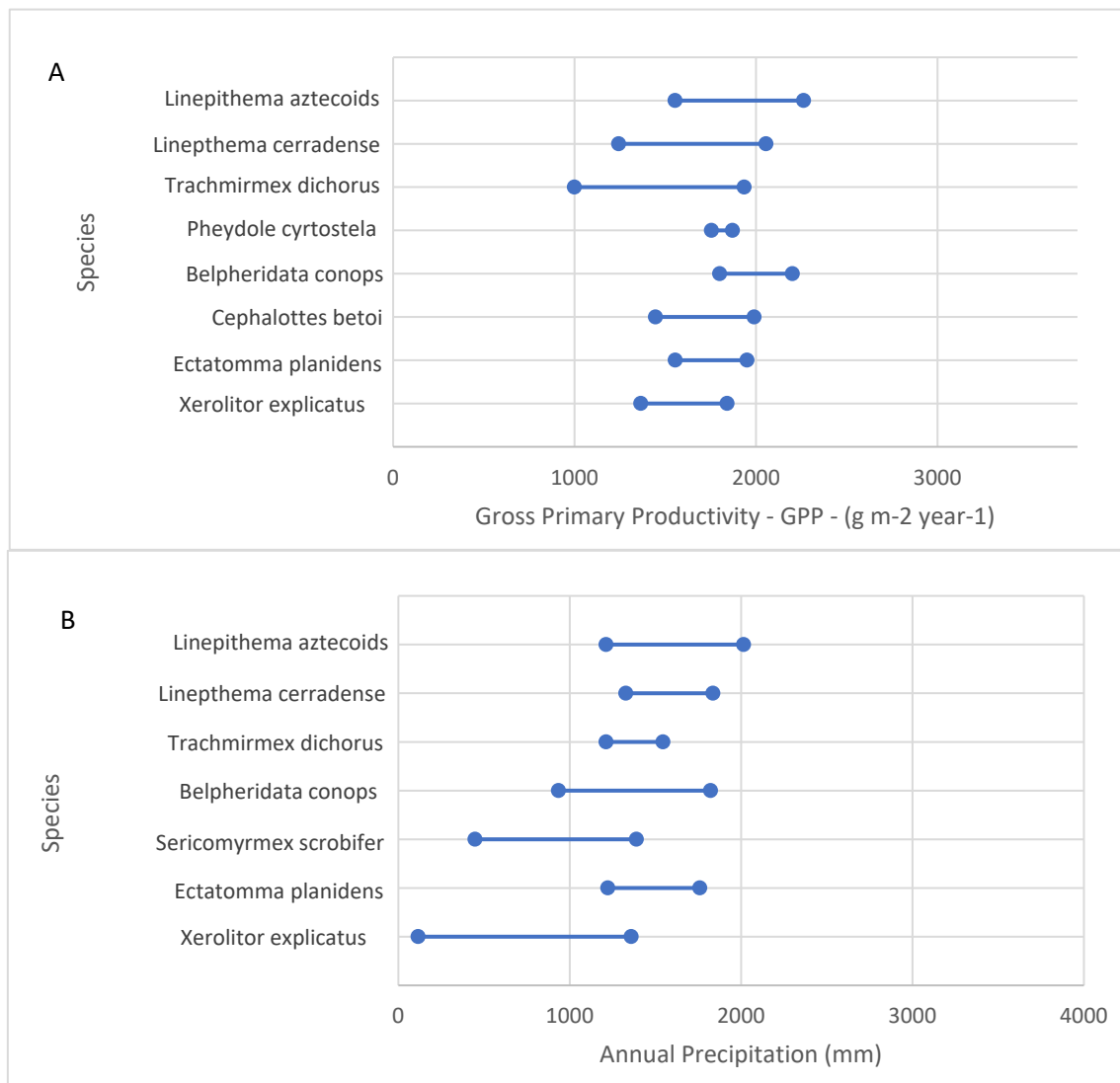
Figures

Figure 1: Intervals with positive weights of evidence for GPP (A) and Annual precipitation (B). Only species whose fitted models included one of those variables are shown. The upper limit of the X axis shows the highest value observed in South America for GPP and, for annual precipitation, the maximum precipitation value observed in 99% of South America territory.
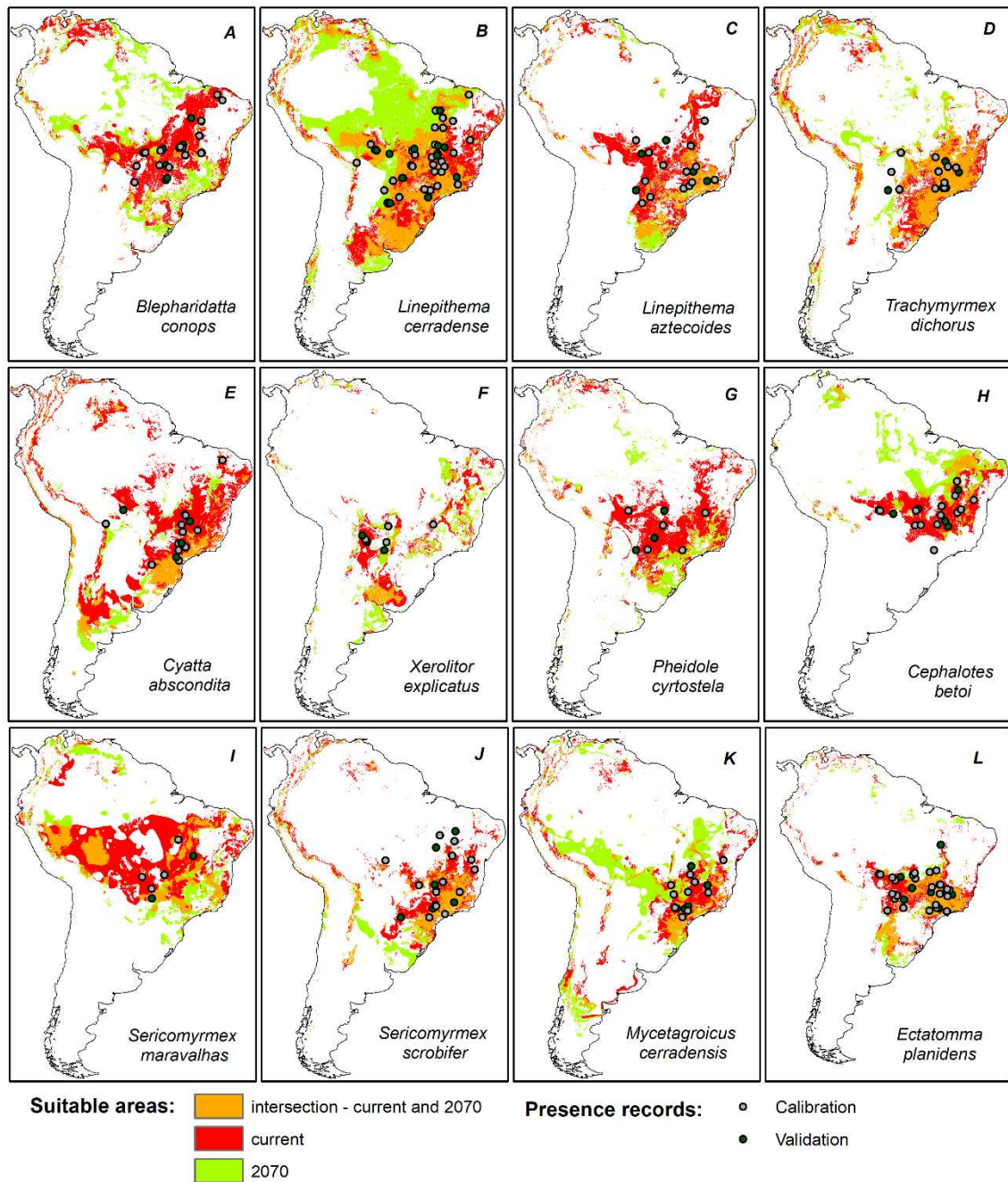
Figure 2. Current and future (2070) expected suitable regions for the occurrence of 12 ant species that occurs exclusively, or predominantly, in savannas or grasslands of central South America.
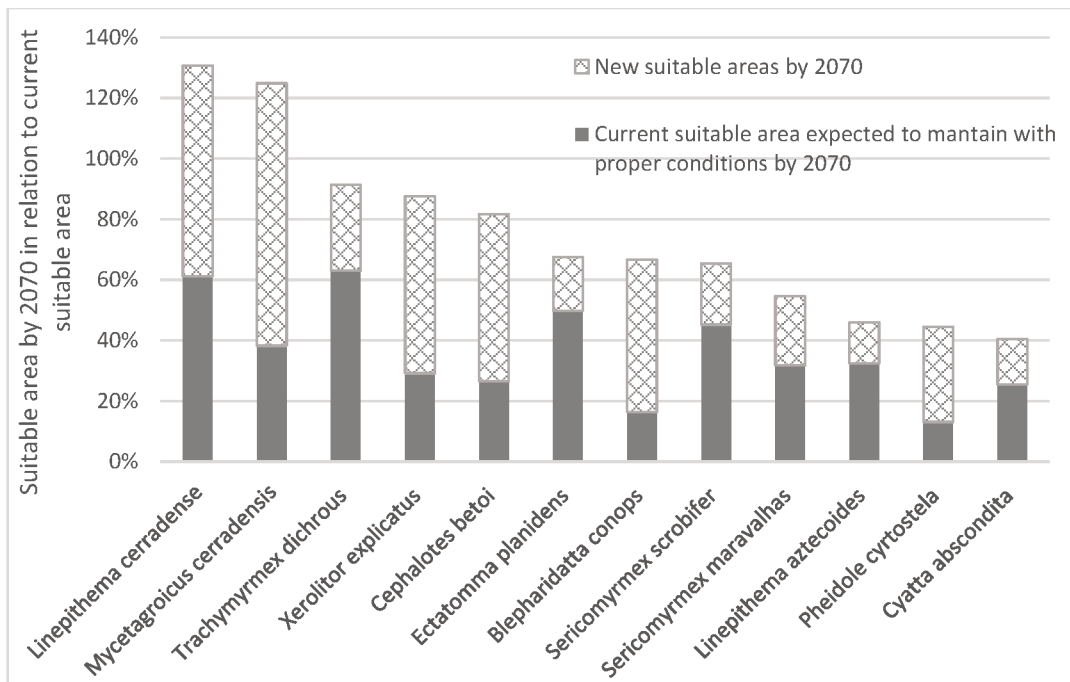
Figure 3: Suitable area by 2070 for the 12 ant species, expressed in percentual rates in relation to current suitable areas. For each species, total percentages by 2070 are divided into areas that are expected to remain with proper conditions by 2070 (gray part of bars) and new zones with suitable conditions by 2070 (hatched portion of bars).
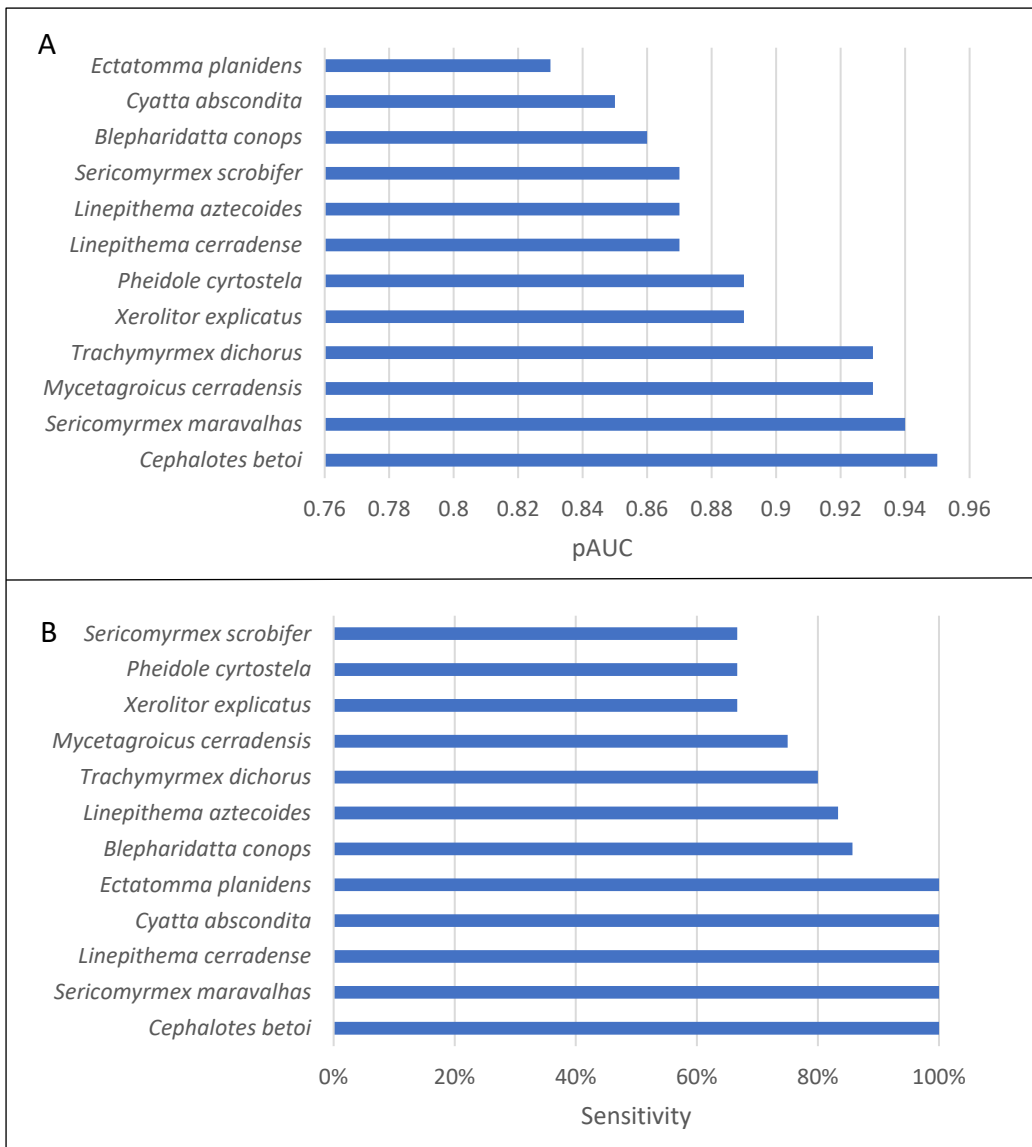
Figure 4: Evaluation of fitted suitability surfaces. Graphs show partial AUC (A) and sensitivity (correct prediction of presences) (B) for each species.
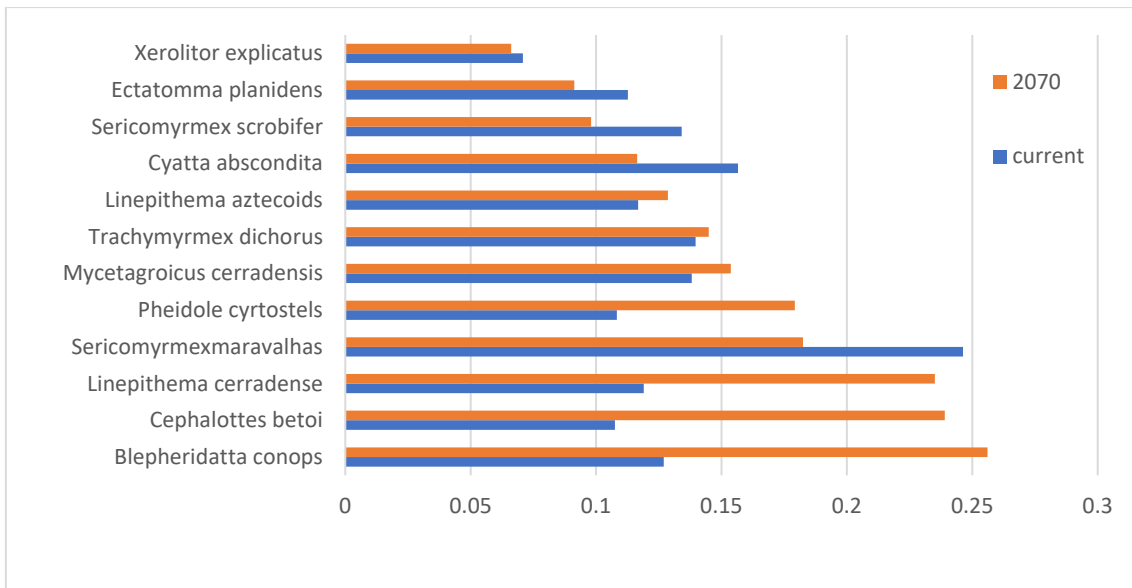
Figure 5: Proportion of the modelled suitable areas of occurrence of each ant species that is represent by protected areas, currently and by 2070.
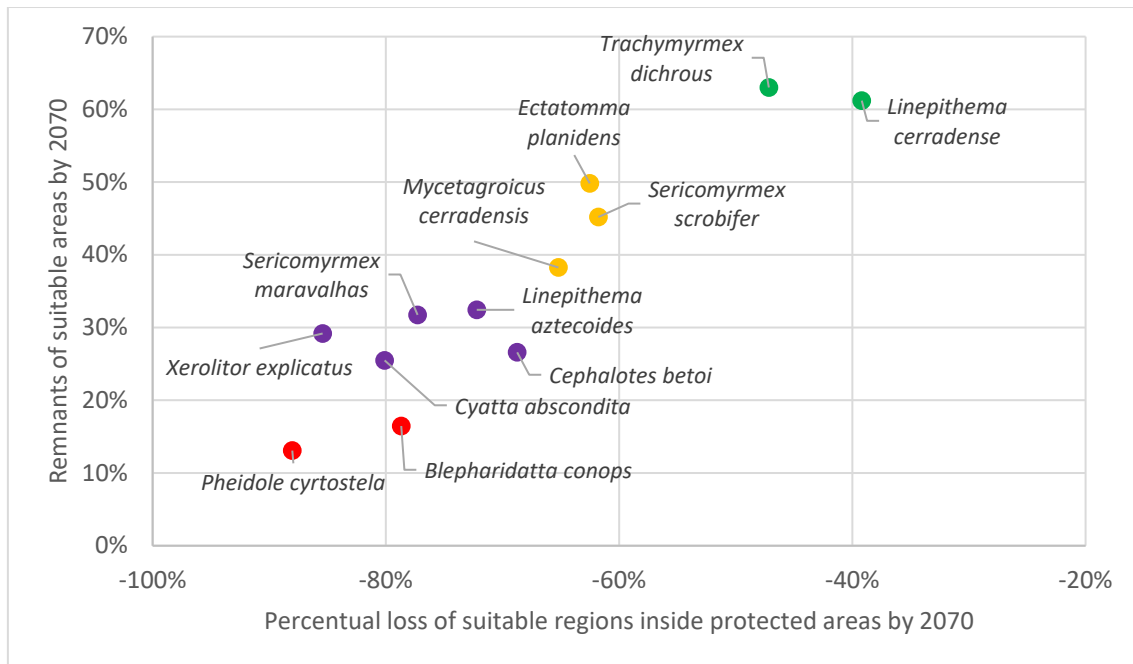
Figure 6: Graph showing vulnerability degree for each species. Vulnerability is based into the relationship between percentage of remnants of suitable areas by 2070 and percentual loss of suitable regions inside Protected Areas (PAs). Vulnerability increases as remnants diminishes and loss of suitable areas inside Protected Areas increases. Species were classified into four vulnerability groups, represented by different colors in the graph. Group I, green group, is the least vulnerable one, comprising species that are expected to maintain more than 60% of current suitable areas by 2070 and have the lowest percentual loss inside PAs. Group II, orange one, contains species which were estimated to have between 39% and 50% of current estimated distributions by 2070 and are expected to loss between 63% and 65% inside protected areas by this same period. Group III, represented by purple dots, includes species whose maintenance percentages varied from 25% to 32% and percentual loss inside protected areas might reach 69% to 85%. Group IV, represented by red dots, is composed of the most threatened species, which are expected to maintain only between 13% and 16% of their current estimated distributions and are expected to loose between 79% and 88% of suitable regions inside PAs.

Tables

Table 1: Species of ants endemic to the savannas, grasslands and/or dry forests of central South America and the number of sites in which they were recorded. Data from Vasconcelos et al. (2017) and from the antmaps.org database.

Species	Number of records
Linepithema cerradense	44
Ectatomma planidens	33
Blepharidatta conops	24
Sericomyrmex scrobifer	22
Cephalotes betoi	22
Linepithema aztecoides	20
Trachymyrmex dichrous	17
Cyatta abscondita	16
Mycetagroicus cerradensis	15
Xerolitor explicatus	9
Pheidole cyrtostela	8
Sericomyrmex maravalhas	6

Table 2: Significant variables included in final Weights of Evidence models of each species. Variables were listed among groups. GPP refers to Gross Primary Productivity. Temperature variables correspond to Annual Mean Temperature (B1), Mean Diurnal Range (B2), Isothermality (B3), Temperature Seasonality (B4), Maximum Temperature of Warmest Month (B5), Minimum Temperature of Coldest Month (B6), Temperature Annual Range (B7), Mean Temperature of Wettest Quarter (B8), Mean Temperature of Warmest Quarter (B10) and Mean Temperature of Coldest Quarter (B11). Rainfall related variables refers to Annual Precipitation (B12), Precipitation of Wettest Month (B13), Precipitation of Driest Quarter (B17), and Precipitation of Coldest Quarter (B19). ALT corresponds elevation and ARG_005 refers to the clay content estimated at 0.05 meter deep.

Species	Variables included in the model, separeted by groups				
	GPP	Temperature	Rainfall	Altitude	Soil
Blepharidatta conops	GPP	B6	B12		
Linepithema cerradense	GPP	B10	B12		
Linepithema aztecoides	GPP	B1	B12		
Cyatta abscondita		B8		ALT	
Xerolitor explicatus	GPP	B2, B10	B12		
Pheidole cyrtostela	GPP	B6, B10	B13		
Cephalotes betoi	GPP	B3, B4, B6	B17		
Sericomyrmex maravalhas		B7, B11	B13		
Sericomyrmex scrobifer		B6, B8	B12	ALT	
Mycetagroicus cerradensis		B2, B5	B13	ALT	
Ectatomma planidens	GPP	B11	B12, B19	ALT	
Trachymyrmex dichorus	GPP	B5	B12, B13	ALT	ARG_005

Conclusões gerais

Esta tese demonstrou a eficácia superior do método Pesos de Evidência em relação ao Maxent quando os dados de entrada são poucos (menos de 30) e abrangem apenas parte da real distribuição da espécie. Este comportamento é essencial tanto na modelagem de espécies raras, ou com poucas informações disponíveis, quanto em situações cujo objetivo é o planejamento da conservação. É importante ressaltar que este trabalho constitui uma primeira avaliação do método Pesos de Evidência na modelagem de distribuição de espécies, portanto o uso em outras espécies e em outros locais de estudo deve ser cautelosa. Idealmente, sugere-se que qualquer modelagem de distribuição de espécies passe pela avaliação de algumas metodologias, aquelas mais adequadas para o conjunto de dados em questão, sendo utilizada, por fim, aquela com maior acerto, o qual deverá ser definido com base em dados não incluídos durante o processo de calibração.

A aplicação do método Pesos de Evidência em formigas do *Cerrado* mostrou que, mesmo sob melhor cenário - considerando que as espécies terão capacidade de migrar e alcançar novas áreas de adequabilidade em 2070 - dez das doze espécies analisadas terão áreas de condições propícias equivalentes a 40% - 90% de suas áreas atuais. Sob uma hipótese mais conservadora, considerando que as espécies não serão capazes de migrar para as novas áreas favoráveis até 2070, esta porcentagem de redução pode ser ainda mais severa, e as áreas futuras poderão corresponder a 13% - 63% das áreas de adequabilidade atuais para as 12 espécies. Em geral, o estudo sugeriu *Linepithema cerradense* como a espécie de menor grau de ameaça. Por outro lado, *Blepheridatta conops* e *Pheidole cyrtostela* podem ser as mais vulneráveis, com menos de 16% da distribuição estimada atual a ser preservada em 2070. Estima-se que, atualmente, as espécies possuam, em média, 13% da área favorável coincidindo com áreas de proteção, valor que pode chegar a 5% em 2070.