bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Land use, macroecology, and the accuracy of the Maximum Entropy Theory of Ecology: A case study of Azorean arthropods

Micah Brush[*,1], Thomas J. Matthews[2,3], Paulo A.V. Borges[3], John Harte[4]

[1]Department of Physics, University of California, Berkeley, Berkeley, CA, United States.
[2]GEES (School of Geography, Earth and Environmental Sciences) and Birmingham Institute of Forest Research, University of Birmingham, Birmingham, United Kingdom.
[3]CE3C – Centre for Ecology, Evolution, and Environmental Changes/Azorean Biodiversity Group and Universidade dos Açores, 9700-042 Angra do Heroísmo, Açores, Portugal.
[4]Energy and Resources Group, University of California, Berkeley, Berkeley, CA, United States.

[*]Corresponding author, micbru@berkeley.edu.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

Human activity and land management practices, in particular land use change, have resulted in the global loss of biodiversity. These types of disturbances affect the shape of macroecological patterns, and analyzing these patterns can provide insights into how ecosystems are affected by land use change. The Maximum Entropy Theory of Ecology (METE) simultaneously predicts many of these patterns using a set of ecological state variables: the number of species, the number of individuals, and the total metabolic rate. The theory's predictions have been shown to be successful across habitats and taxa in undisturbed natural ecosystems, although previous tests of METE in relation to disturbance have focused primarily on systems where the state variables are changing relatively quickly. Here, we assess predictions of METE applied to a different type of disturbance: land use change. We use METE to simultaneously predict the species abundance distribution (SAD), the metabolic rate distribution of individuals (MRDI), and the species–area relationship (SAR) and compare these predictions to arthropod data from 96 sites at Terceira Island in the Azores archipelago across four different land uses of increasing management intensity: 1. native forest, 2. exotic forest, 3. semi-natural pasture, and 4. intensive pasture. Across these patterns, we find that the forest habitats are the best fit by METE predictions, while the semi-natural pasture consistently provided the worst fit. The intensive pasture is intermediately well fit for the SAD and MRDI, and comparatively well fit for the SAR, though the residuals are not normally distributed. The direction of failure of the METE predictions at the pasture sites is likely due to the hyper-dominance of introduced spider species present there. We hypothesize that the particularly poor fit for the semi-natural pasture is due to the mix of arthropod communities out of equilibrium and the changing management practices throughout the year, leading to greater heterogeneity in composition and complex dynamics that violate METE's assumption of static state variables. The comparative better fit for the intensive pasture could then result from more homogeneous arthropod communities that are well adapted to intensive management, and thus whose state variables are less in flux.

Key words: Land use, macroecology, Maximum Entropy Theory of Ecology, species abundance distribution, metabolic rate distribution, species–area relationship, Azores, arthropods, disturbance

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 Introduction

Human management of land, and particularly the act of land use change, is a primary driver of ecological disturbance around the world (Foley et al. 2005; Pereira et al. 2012; Klein Goldewijk et al. 2017). Land use changes have large effects on landscape heterogeneity, and have resulted in the broad scale loss and fragmentation of natural habitats, resulting in a mosaic of habitat types in most landscapes (Fahrig 2003; Fischer and Lindenmayer 2007; Cardoso et al. 2009; Fahrig 2019). This type of human driven disturbance has resulted in global biodiversity loss (Martins et al. 2014; Pimm et al. 2014; Newbold et al. 2015; Maxwell et al. 2016; Newbold et al. 2018). On oceanic islands, the conversion of native vegetation to forestry (managed forest plantations – monocultures of fast growing trees), agricultural, and pasture land has had particularly severe impacts on the native biota due to the small-scale nature of islands, the sensitivity and small ranges of many island endemics, and the fact that land-use change on islands has often been accompanied by the spread of exotic species (Gillespie and Roderick 2002; Borges et al. 2006; Whittaker and Fernández-Palacios 2007; Gillespie et al. 2008; Whittaker et al. 2017).

Disturbance as a result of human activity can be expected to impact the form of macroecological patterns (Gray et al. 1979; Hill and Hamer 1998; Dornelas et al. 2009; Newman 2019), including through land use changes (Simons et al. 2015; Xu et al. 2019). An effective method for analyzing the impacts of disturbance on biodiversity is comparing empirical patterns to theoretically expected shapes (Kempton and Taylor 1974; Carey et al. 2006; Supp et al. 2012; Matthews and Whittaker 2015; Franzman et al. 2021) and interpreting the deviations in the context of the type of disturbance. In order to interpret macroecological patterns in this way, we require a theoretical prediction for what these different patterns should look like in an ecosystem or habitat that has not been disturbed or managed.

The Maximum Entropy Theory of Ecology (METE) provides one such tool. METE predicts numerous macroecological patterns simultaneously using the principle of maximizing information entropy (Harte et al. 2008; Harte 2011; Harte and Newman 2014; Brummer and Newman 2019). METE is characterized by three so-called state variables that are used to constrain the predicted distributions for a given ecosystem or habitat: the species richness S_0, the number of individuals N_0 and the total metabolic rate E_0. To make spatial predictions, METE also requires the total area of the site A_0. METE has been found to well describe empirical patterns across diverse taxa and habitats (Harte 2011; White et al. 2012; Xiao et al. 2015). However, there is increasing evidence that these predictions perform less well in disturbed ecosystems (Carey et al. 2006; Rominger et al. 2016; Newman et al. 2020; Franzman et al. 2021; Harte et al. 2021). In the context of METE, most disturbance has been characterized by ecosystems with rapidly changing state variables, as METE seems to well describe ecosystems where state variables are relatively constant in time. This means

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

that deviations from METE predictions may be capturing information about the rate of change of state variables.

The effects of land use change on the deviation of macroecological patterns from METE predictions has not yet been explored. Given that METE predictions appear to perform better in pristine ecosystems, we expect that anthropogenic land uses that introduce significant disturbance should result in patterns that deviate from them in meaningful ways (Harte et al. 2021). How well the data fit METE across land use types can thus provide insights about how different land uses affect these large scale patterns, and by extension how land use change affects biodiversity.

Here, we investigate how land use change affects several patterns predicted by METE with arthropod data from Terceira Island in the Azores archipelago (Portugal). The Azores are an isolated island chain in the Atlantic Ocean that have been populated for about 600 years (Norder et al. 2020). In this time, they have undergone a large change from largely undisturbed natural forest to mixed land uses, including managed forest plantations (usually monocultures of fast growing trees) and intensive pasture land (Cardoso et al. 2009).

Previous work based on the analysis of Azorean arthropod data has shown that a variety of macroecological patterns, such as the species abundance distribution and functional trait composition, vary as a function of land use (Fattorini et al. 2016; Borda-de-Água et al. 2017; Rigal et al. 2018). Thus, it represents an ideal system to test how land use changes affect macroecological patterns and their deviation from METE predictions. In addition, the dataset investigated is unique since it incorporates detailed body-size measurements that can be translated to body mass data for an important fraction of sampled arthropod taxa.

We test three predictions of METE simultaneously: the species abundance distribution (SAD), the metabolic energy rate distribution of individuals (MRDI), and the species–area relationship (SAR). See Table 1 for more information about these patterns and their METE predicted forms. Analyzing multiple patterns simultaneously avoids treating any individual pattern in isolation, especially since single patterns can often be predicted from many different underlying theories (McGill et al. 2007). We compare all three of these patterns with arthropod data across four land use types on Terceira Island, and analyze the deviations from the predicted patterns for information about how disturbance is affecting species community assembly. We predict that METE predictions will fit the data better for less intensively managed land uses, with deviations linearly increasing with management intensity.

Additionally, many arthropod species have been introduced to the archipelago by humans (Borges et al. 2010). These exotic species have changed the ecological landscape (Florencio et al. 2013) and have been found to have a different functional trait composition from the indigenous species (Rigal et al. 2018). However, some studies have found that these species appear to be integrated in these ecosystems, perhaps by replacing lost indigenous species or filling empty niche

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

space (Gaston et al. 2006; Rigal et al. 2013). We might therefore expect that METE predictions would perform better when indigenous and introduced species are analyzed together. To test this, we analyze indigenous and exotic species separately, in addition to our analyses with all species together.

2 Methods

2.1 Study area and arthropod data

The Azores Islands are an isolated island chain in the Atlantic Ocean of volcanic origin. All of the data analyzed here come from Terceira Island, which before human colonization was almost entirely forested but now comprises a mix of land uses. The four major land uses, ranked in increasing order of management intensity, are 1. native forest, 2. exotic forest, 3. semi-natural pasture, and 4. intensive pasture (Cardoso et al. 2009; Rigal et al. 2018). These land uses comprise about 87% of the total island area, which is broken down by land use in Table 2 (Cardoso et al. 2009). Figure S1.1 in Rigal et al. (2018) shows a land use distribution map of Terceira Island with more specific spatial information.

The native forest is made up of perennial trees and shrubs adapted to a hyper-humid Atlantic climate, and is now restricted to elevations above 500 m above sea level (a.s.l.) and dominated by *Juniperus-Ilex* forests and *Juniperus* woodlands (Elias et al. 2016). Exotic plantations of the fast growing tree *Cryptomeria japonica* were planted after the Second World War to reforest large areas of previous native forest that were destroyed in the previous decades for fuel. These plantations are dense and almost no understory is present. Semi-natural pastures are located around 400-600 m a.s.l., have a mixture of native and exotic herbs and grasses, and are mostly grazed in the spring and summer with low cattle density. Intensive pastures are located between 100-500 m a.s.l. and are grazed every three weeks (and sometimes up to every 12 days in the summer) with high cattle density.

The arthropod samples were collected using pitfall traps across 96 sites. Each of the sites has a single 150 m transect with 30 pitfall traps spaced out at 5 m intervals: 15 traps filled with approximately 60 mL of a non-attractive solution (anti-freeze liquid) with a small proportion of ethylene glycol, and 15 traps with the same volume of a general attractive solution (Turquin), which was made of dark beer and some preservatives. All data were collected over summers on Terceira Island over the period from 1997-2009 (for more details, see Borges et al. 2005; Cardoso et al. 2009; Rigal et al. 2018).

Table 2 shows the number of sites for each land use, along with the total and median number of species S_0 and individuals N_0 across all transects in that land use. The number of indigenous and

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

exotic species or individuals are shown in parentheses. Indigenous species are defined as those that are endemic (occur only in the Azores) or native (appear in the Azores Islands and other nearby archipelagos and mainland). Exotic species are those believed to have been introduced by humans following human colonization of the archipelago in the 15th century, and therefore we will also refer to them as introduced species (Borges et al. 2010). Unidentified species that share a genus, subfamily, or family with other species present in the archipelago are put into the same colonization category as those species (Borges et al. 2010; Florencio et al. 2013). There are four remaining species that are not identified as indigenous or exotic that constitute 11 individuals. These species are removed from the analysis when indigenous and introduced species are analyzed separately. Across all land uses, there are a total of 271 species and 46 250 individuals, with 126 indigenous species and 14 950 indigenous individuals and 141 exotic species and 31 289 exotic individuals. However, juvenile individuals are excluded from the MRDI analysis, leaving 226 species and 36 269 individuals, and there are less data available with the spatial resolution needed for the SAR analysis, with 228 species and 34 282 individuals.

Body length measurements of individuals from the different species are obtained as described in Rigal et al. (2018), and average body length values are used here. For 26 Araneae species, we use updated body length measurements taken from a new database of Macaronesian spider traits (Macías-Hernández et al. 2020). Body lengths are then converted to body mass values using empirical scaling equations, as detailed in Appendix S3 (Hódar 1996; Baumgärtner and Rothhaupt 2003; Wardhaugh 2013). Additionally, body mass variation within a single species is reintroduced by assuming a normal distribution for intraspecific body size (Gouws et al. 2011), and then obtaining parameters for this distribution by relating the mean and variance of body mass for several spider (Macías-Hernández et al. 2020) and beetle (Terzopoulou et al. 2015) species. We then simulate body masses for all individuals using the parameters obtained for beetles for all orders except for spiders, where we use the parameters obtained for them. Since Coleoptera and Araneae are the two most common orders in the dataset, differences among other orders should not overly impact the analysis. For more information on how we introduce intraspecific body mass variation, see Appendix S4.

2.2 METE Review

METE predicts many different macroecological patterns simultaneously by maximizing Shannon information entropy given a set of constraints (Harte et al. 2008; Harte 2011; Harte and Newman 2014; Brummer and Newman 2019). Its core distribution is the ecological structure function $R(n, \varepsilon|S_0, N_0, E_0)$, which is a joint distribution over abundance n and metabolic rate ε given the number of species S_0, the number of individuals N_0, and the total metabolic rate E_0. Thus, $Rd\varepsilon$ is

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the probability that an individual picked at random from all species with abundance n has metabolic rate between ε and $\varepsilon + d\varepsilon$. Note that n is discrete, and ε is continuous. In practice, we scale the metabolic rate such that the smallest metabolic rate in the ecosystem has $\varepsilon = 1$.

We use the method of Lagrange multipliers to maximize the information entropy $\sum_n \int_\varepsilon d\varepsilon R \log(R)$ given the following constraints:

$$\begin{aligned} \frac{N_0}{S_0} &= \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \; nR(n,\varepsilon) \\ \frac{E_0}{S_0} &= \sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \; n\varepsilon R(n,\varepsilon). \end{aligned} \tag{1}$$

We additionally require the distribution R to be normalized such that $\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \; R(n,\varepsilon) = 1$. The solution for the ecological structure function is

$$R(n,\varepsilon|S_0,N_0,E_0) = \frac{\exp\left(-\lambda_1 n - \lambda_2 n\varepsilon\right)}{Z}, \tag{2}$$

where the Lagrange multipliers λ_1 and λ_2 are solved from the constraints, and the normalization Z is calculated as $\sum_{n=1}^{N_0} \int_{\varepsilon=1}^{E_0} d\varepsilon \; \exp\left(-\lambda_1 n - \lambda_2 n\varepsilon\right)$. To a very good approximation given typical empirical values for S_0, N_0, and E_0 (Harte 2011; Brummer and Newman 2019), $\lambda_2 = S_0/(E_0 - N_0)$, and λ_1 can then be solved from

$$\frac{N_0}{S_0} = \frac{\sum_{n=1}^{N_0} e^{-\beta n}}{\sum_{n=1}^{N_0} e^{-\beta n}/n}, \tag{3}$$

where $\beta = \lambda_1 + \lambda_2$.

The distribution R can then be used to derive other macroecological distributions that can be compared to data. We here show derivations of the relevant distributions. The results are also summarized in Table 1.

2.2.1 Species abundance distribution

We obtain the METE SAD prediction $\Phi(n)$ by integrating the structure function over ε. The METE prediction is equivalent to the max likelihood prediction for the log series (White et al. 2012, Appendix A). This prediction assumes that the number of species is large enough that we can ignore certain terms, which eliminates any dependence on E_0. The resulting prediction is the log series distribution

$$\Phi(n|S_0,N_0) = \frac{e^{-\beta n}}{n \log\left(1/\left(1-e^{-\beta}\right)\right)} \tag{4}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.2.2 Metabolic rate distribution of individuals

The METE MRDI prediction $\Psi(\varepsilon)$ is obtained by summing the structure function multiplied by n over n and correcting the normalization,

$$\Psi(\varepsilon) = \frac{S_0}{N_0} \sum_{n=1}^{N_0} nR(n, \varepsilon). \tag{5}$$

This sum gives

$$\Psi(\varepsilon|S_0, N_0, E_0) = \lambda_2 \left(e^{\beta} - 1\right) \frac{e^{-\gamma}}{(1 - e^{-\gamma})^2}, \tag{6}$$

where $\gamma = \lambda_0 + \lambda_1 \varepsilon$. Note that we use a slightly different form for Ψ compared to Eq. 7.33 in Harte (2011), where β has been replaced with $e^{\beta} - 1$. The normalization of $\Psi(\varepsilon)$ in Eq. 6 is significantly better, as $\int_{\varepsilon} d\varepsilon\ \Psi(\varepsilon)$ is much closer to 1. This form still allows the cumulative distribution function and the rank ordered distribution to be solved analytically, and is numerically very similar to the full expression without any approximations, even for analyses at the transect level.

2.2.3 Species–area relationship

The SAR can be predicted by combining the SAD with the species-level spatial abundance distribution (SSAD) $\Pi(n|A, A_0, n_0)$, which predicts the number of individuals present in an area A given n_0 individuals of that species in a larger area A_0. The number of species at a given scale A can be predicted by multiplying the SAD by the probability that a species with abundance n_0 is present at that scale and then summing over n_0,

$$S(A) = \sum_{n_0=1}^{N_0} \Phi(n_0) \left(1 - \Pi(0|A, A_0, n_0)\right). \tag{7}$$

The SSAD can also be predicted by maximizing information entropy given the constraint $\sum_{n=0}^{n_0} n\Pi(n) = n_0 A/A_0$. The solution corresponds to the finite negative binomial distribution (Conlisk et al. 2007; Zillio and He 2010)

$$\Pi(n|A, A_0, n_0) = \frac{\binom{n+k-1}{n}\binom{n_0-n+kA_0/A-k-1}{n_0-n}}{\binom{n_0+kA_0/A-1}{n_0}}, \tag{8}$$

with aggregation parameter $k = 1$ (Harte 2011; Wilber et al. 2015).

METE also predicts that all nested SARs will collapse onto a single universal curve when plotted as the slope of the SAR z versus $D = \log(N_0/S_0)$, a scale parameter (Harte 2011; Wilber et al. 2015).

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2.3 Comparing METE predictions with data

We divide the data by land use and compare the observations to the predictions of METE. Because we have multiple transects for each land use, we can either compare our results by aggregating across all transects in one land use category, or at the individual transect level by treating each transect as a replicate. We primarily analyze our results at the individual transect level. This is because METE makes predictions within a single community, and aggregating data over several locations may create a mismatch between the theory predictions and the data. Additionally, the number of species observed will not be the same across many small patches compared to a single large patch of the same area. Despite these issues, we do find mostly similar results when aggregating the data by land use (Appendix S1).

Our sampled arthropod data are separated into juvenile and adult individuals. For the SAD and SAR analysis, we treat these together as a single dataset that accurately captures all ground dwelling arthropods. Note that some species will have adult forms that will not be captured by pitfall traps. However, in some sense the arthropods that are not captured by the pitfall traps are not part of the same ecosystem of ground dwelling arthropods (Borges et al. 2005; Gaspar et al. 2008). These individuals likely interact with different species and habitats, and thus do not necessarily need to be included for METE to make accurate predictions about the composition of the study ecosystem. For the predictions of the metabolic rate distribution of individuals (MRDI), we are only able to use adult arthropods as the scaling relationships used to calculate the empirical metabolic rates are only available for adult individuals (see Appendix S3).

Finally, we categorize species as indigenous (native or endemic) or exotic (likely introduced by humans). In addition to our primary analysis treating the species together as one community, we analyze these groups independently for all patterns.

2.3.1 Species abundance distribution (SAD)

There are many existing methods for comparing empirical SADs to data; however, there is not a single optimum goodness of fit metric (Connolly and Dornelas 2011; Matthews and Whittaker 2014). In general, binning should be avoided as the shape of the distribution depends on the binning interval, and given the sparse data each observed bin is unlikely to have a sufficiently large expectation to have a meaningful χ^2 test (Williamson and Gaston 2005; Gray et al. 2006; Ulrich et al. 2010). Log-likelihood based methods are also inappropriate in this case as we are not trying to determine a preferred model and instead are looking for a goodness of fit test.

Given this, we use the mean least squares of the rank ordered natural log of abundance as our primary goodness of fit metric to compare the METE predictions to data. Mathematically, this means we take $\sum_{i=1}^{S_0} \left(\log\left(n_{i,\text{observed}}\right) - \log\left(n_{i,\text{predicted}}\right)\right)^2 / S_0$, where n_i is the abundance of the

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

species with rank i, and we take the mean over all S_0 ranks. Although Matthews and Whittaker (2014) note that mean least squares of rank ordered data violates some underlying statistical assumptions, namely that the data points are not statistically independent (see also Connolly and Dornelas 2011), we still find it preferable to their recommended method of using a parametric bootstrap. For many test statistics this method scales with the number of points, making comparisons between land use types challenging. As that is our primary objective here, we prefer to use the mean least squares.

To ensure our results are robust to our choice of goodness of fit metric, we additionally performed Kolmogorov-Smirnov (KS) tests for each transect and obtained the two-sided test statistic D_{KS} when the empirical cumulative distribution function (CDF) is compared to the METE predicted CDF. These results can be found in Appendix S2, and are very similar to the results obtained by mean least squares. For the KS test for the SAD, we used the R package DGOF which implements the KS test for discrete distributions (Arnold and Emerson 2011).

2.3.2 Metabolic rate distribution of individuals (MRDI)

We first reintroduce intraspecific body mass variation as described above and in Appendix S4. For adult arthropods, we then use metabolic scaling to convert body mass data to metabolic rate. We assume that $\varepsilon \propto m^{3/4}$, where m is the body mass. The metabolic rates are then scaled such that the smallest $\varepsilon = 1$. We then rank order the data in order to compare to the METE prediction.

As with the SAD, we primarily use the mean least squares of the rank ordered natural log of metabolic rates as a goodness of fit metric for each land use. In this case, that means $\sum_{i=1}^{N_0} \left(\log\left(\varepsilon_{i,\text{observed}}\right) - \log\left(\varepsilon_{i,\text{predicted}}\right)\right)^2 / N_0$, where ε_i is the metabolic rate of the individual with rank i and the mean is now over the number of individuals N_0. Alternative options for this comparison would be R^2 as defined in eg. Xiao et al. (2015), or to bin the data and use a χ^2 test. However, binning relies on a large number of points per bin that are not available here at the individual transect level, and the results can depend on the bin width. In order to test the robustness of our goodness of fit metric we additionally perform Kolmogorov-Smirnov goodness of fit tests for the empirical CDF compared to the METE predicted CDF. Again, we obtain similar results to the mean least squared analysis (Appendix S2).

2.3.3 Species–area relationship (SAR)

Along each transect, there are 30 individual pitfall traps arranged linearly. For each transect individually, we compare the resulting empirical SAR to the METE prediction by first averaging the number of species at different scales. We choose scales of 1, 2, 3, 5, 6, 10, 15, and 30 traps. These relative scales were chosen as they use all of the data available at every scale, or put another

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

way, these numbers are all factors of 30. This is slightly different than has been done in other comparisons, which use a number of cells that is a factor of two and divide repeatedly in half (eg. Franzman et al. 2021).

To compare these average numbers of species to the METE prediction, we use the functions for the log series SAD $\phi(n)$ and the finite negative binomial SSAD $\Pi(n)$ from the `macroeco` software package (Kitzes et al. 2015; Kitzes and Wilber 2016), which we have adapted for use with Python 3.0.

There are many different methods for comparing the predicted SAR to data. We could fix the number of species and individuals at the largest scale and predict the number of species at every smaller scale from this anchor scale, or we could compare the predicted slope z of the SAR against the scale parameter $D = \log(N_0/S_0)$, which as noted above collapses SARs onto a single, universal curve (Harte et al. 2009; Wilber et al. 2015). Harte (2011) provides an equation for z given the empirical number of species and individuals at a given scale, assuming that we bisect the plot in two. We use a similar approach, in that we use the number of species and individuals at a given scale to predict the number of species at the next smallest scale in consideration and then use that to predict a slope. Mathematically, we write

$$z_i = \frac{\log(S_i/S_{i-1})}{\log(A_i/A_{i-1})}, \tag{9}$$

where i indexes the scale. In the list above then, $i = 1$ corresponds to the average number of species at the scale of 1 cell, and $i = 8$ corresponds to the number of species in all 30 cells.

Given we are predicting the number of species directly using this approach, we can also compare the SAR directly (see Appendix S5). However, given the collapse onto a single curve in the $z - D$ plots we prefer that comparison. We obtain the empirical slope by comparing the number of species at the scale being considered to the number of species at the next smallest scale. This is similar to the theoretical prediction, but now the number of species at the smaller scale is also empirical. This method allows us to compare slopes at all except the smallest scale, as we do not have a smaller scale with which to make the empirical slope prediction.

We therefore have seven data points to make the comparison and again use the mean least squares between the METE prediction and the empirical data, here as $\sum_{i=2}^{8} \left(z_{i,\text{observed}} - z_{i,\text{predicted}}\right)^2 /7$, where i indexes the scale and we do not compare at the smallest scale $i = 1$. Note that for many transects we will have fewer than seven data points as we additionally only use scales where the empirical average for the number of species is greater than four ($S_0 > 4$). This is because several METE simplifications break down for small S_0, including the fact that we can ignore E_0 and derive β from only N_0 and S_0. This means that transects with lower abundance will have fewer than seven points of comparison.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

3 Results

3.1 Species abundance distribution (SAD)

The circle markers in Fig. 1 show the mean of the mean least squared error over transects with its standard error at each land use for the SAD. We find that the semi-natural pasture is particularly poorly described by METE, and the native forest is the best fit. The results for the exotic forest and intensive pasture are fairly similar and intermediate between the native forest and the semi-natural pasture. The standard deviation of the mean is the lowest for the native forest sites, and the highest for the semi-natural pasture. Similar results for the KS test statistic D_{KS} across transects are shown in Appendix S2.

Plots of the SAD at each transect are shown in Appendix S6. To combine all of the SADs onto a single plot, we plot the residuals of $\log_{10}$(abundance) in Fig. 2, where the x-axis has been scaled by the number of species to facilitate comparison. The residuals are calculated as $\log_{10}(\text{abundance}_{\text{observed}})$ - $\log_{10}(\text{abundance}_{\text{predicted}})$. Each line in this plot represents the SAD at a single site. In the semi-natural pasture in particular, we see that METE consistently under predicts the abundance of the most abundant species and over predicts the abundance of the species of intermediate rank. We can see a similar pattern across land uses where the residual of the most abundant species tends to be positive, though it is most prevalent at the pasture sites and least common at the native forest sites. Across all sites, METE generally under predicts the number of singletons, though this is again less common in the native forest sites.

Results of aggregating over transects rather than treating them as replicates can be found in Appendix S1. In that case, the forest sites are again better fit than the pasture sites, but the ranking is slightly different. The overall goodness of fit is also worse, which is in line with our expectation that METE predictions are more accurate for individual transects.

3.2 Metabolic rate distribution of individuals (MRDI)

The triangle markers in Fig. 1 show the mean and its standard error for the mean least squared error across transects for the MRDI. This metric shows that the forest sites are better fit by METE than the pasture sites, and the semi-natural pasture sites are again particularly poorly fit by METE.

The rank ordered plots for the MRDIs at each transect are plotted in Appendix S7, and Fig. 3 shows the residuals, calculated as observed minus predicted, from the rank ordered metabolic rates at each transect on a single plot. Each line here represents a single site at that land use. Across land uses, we see that there are often far more individuals around a single metabolic rate than predicted by METE at any given site. This appears as long lines of similar slope in the residuals. See particularly the pasture sites, where this pattern in the residuals is especially common.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The results are again similar if analyzed using the KS test statistic (Appendix S2) or at the community level (Appendix S1, though here the fit at the intensive pasture is much worse).

3.3 Species–area relationship (SAR)

The square markers in Fig. 1 show the mean of the mean least squared error over transects with its standard error at each land use for the slope z. Note the different y-axis scale compared to the SAD and MRDI. Here, the mean least squared error for each transect is averaged over the number of scales where the empirical $S_0 > 4$. The results here are comparable across land use types, though the semi-natural pasture appears to be the worse fit.

Figure 4 compares the SAR data for each site, organized by land use. Each point here represents a single transect at a single scale $D = \log(N_0/S_0)$, and the lines are the corresponding METE predictions. There is a large amount of scatter in these plots across land use types. The residuals, again calculated as observed minus predicted, are shown below the plot for the corresponding land use. Looking at the residuals, we see that METE tends to under predict the slope at larger scales. We can see this pattern across sites in Fig. 5, which uses the scale-collapse of the $z-D$ relationship to display all of the sites together on a single plot.

The results are similar if we analyze the predicted number of species at each scale rather than the slope (Appendix S5).

3.4 Indigenous and exotic species

We analyze all of the macroecological patterns independently for species classified as indigenous and exotic at each land use. Figure 6 shows the mean and associated standard error of mean least squares across transects, separated by species that are indigenous and introduced, at all four different land uses. The number of sites where we are able to separate the individuals in this way is limited for the MRDI as adults of all species are not present at each site, and for the SAR as we again only use scales where $S_0 > 4$. Note the difference in scale between this figure and Fig. 1.

For the SAD, we find that the largest difference between the indigenous and introduced species is at the semi-natural pasture sites, where the introduced species fit quite poorly compared to the indigenous species. Across other land uses indigenous and introduced species are comparably well fit, though the introduced species fit slightly better at the exotic forest sites. These fits are also generally comparable to the combined fits, except that the fit of the indigenous species at the exotic forest and the introduced species at the semi-natural pasture are worse (Fig. 1).

For the MRDI, we find a large difference at the native forest sites, where the introduced species again fit poorly compared to the indigenous species. We see a slight difference at the intensive pasture site, where the introduced species fit slightly better. We also find a very different trend

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

overall when the species are analyzed in this way, in that the semi-natural pasture is no longer the worst fit to the METE predictions. For the indigenous species the goodness of fit is similar between the native forest and the semi-natural pasture, and is worse but again similar at the exotic forest and intensive pasture sites. For the introduced species, the semi-natural pasture has the best fit, though the intensive pasture is close, and at both forest sites the fit is much worse than when the species are analyzed together. Additionally, the fits for the introduced species in the native forest and the introduced and indigenous species at the exotic forest are all worse than the fit when indigenous and introduced species are combined (Fig. 1).

For the SAR, the larger error and smaller differences make the analysis more difficult, though we do seem to see a trend of decreasing mean least squares with land use intensity for the introduced species, and increasing mean least squares for the indigenous species. This means that for the first three land uses, the indigenous species are better fit, and only at the intensive pasture are the introduced species better fit. The fit for the indigenous species across all land uses except the intensive pasture is comparable to that of the combined fit (Fig. 1).

4 Discussion

In this study we have compared, for three commonly studied macroecological patterns (the SAD, MRDI, and SAR), the predictions of METE to the empirical patterns of Azorean arthropods across four land use types of varying management intensity. Overall we find that, when all species (indigenous and exotic) are considered together, METE provides a reasonable approximation to the patterns in the native forest and exotic forest sites, as predicted. In contrast, METE provides the worst fit to the semi-natural pasture sites. Interestingly, METE provides a relatively better fit, similar to that for the exotic forest sites, to the most intensively managed land use sites, the intensive pasture. We now discuss the results for each of the three patterns in turn. Unless otherwise noted, we discuss the results when indigenous and introduced species are combined.

4.1 Species abundance distribution (SAD)

The most distinctive pattern in the SAD residuals in Fig. 2 is the consistent under prediction of the most abundant species, particularly at the pasture sites. In other words, these sites have a few very abundant species that are much more abundant than predicted by METE. This result is comparable to the findings of Simons et al. (2015) and Xu et al. (2019), who found that increased grazing intensity (or in the case of Simons et al. (2015), land use intensity as measured by a combination of fertilization, mowing, and grazing) led to a reduction in evenness and an increase in the dominance of the most abundant species (ie. hyper-dominance sensu Hubbell 2013). However, note that in

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

our case the intensive pasture, which had the highest grazing and land use intensity, fit better than the semi-natural pasture. We attribute the dominance of a few very abundant species to small-bodied, highly dispersive, mostly introduced spider species, which were found to be very prevalent at sites with high land use intensity (Borges and Wunderlich 2008; Rigal et al. 2018). This hyper-dominance is an important characteristic of numerous communities (Hubbell 2013; Steege et al. 2013) and should be investigated further as it can be related to species diversification (McGill et al. 2019).

A previous study using Azorean arthropod data found that species dispersal ability affects the form of the SAD (Borda-de-Água et al. 2017). It was found that when species are grouped together by dispersal ability, the corresponding Preston plots (number of species versus $\log_2$(abundance)) are steep for high dispersal species and develop an intermediate mode for lower dispersal species. When rank ordered, this results in SADs that are steeper at low rank for high dispersal species, and less steep at low rank for lower dispersal species. We see very steep rank ordered SADs in the pasture sites in Fig. 2 and in Appendix S6, which is thus likely a result of the highly dispersive spider species. These species are mostly introduced, and thus have not yet evolved to have reduced dispersal ability as they have only been present on the island for a short period of time relative to evolutionary time scales (Borges and Wunderlich 2008).

These mostly introduced spider species additionally have multiple generations per year, which allows them to recolonize the pasture sites from the nearby surrounding forest after disturbance due to land management or cattle grazing (see Rigal et al. 2018, Figure S1.1 for a map). This recolonization is particularly relevant for the semi-natural pasture sites, where cattle grazing and fertilization is seasonal (Borges and Brown 2004). The particularly high abundance of these spiders species in these data may also be related to the fact that these data were collected in the summers, when the semi-natural pasture sites are likely to be subject to cattle grazing.

Another pattern observed in the residuals is that METE tends to under predict the abundance of the intermediate rank species, again particularly in the pasture sites. However, given that METE is constrained such that the number of individuals must be N_0, this is likely a consequence of these highly abundant introduced spider species. If the METE SAD under predicts the abundance at some rank, it must over predict the abundance at another rank such that this constraint is satisfied.

Similarly here, METE tends to under predict the number of singletons at most sites, except in some cases for the native forest sites where we see that METE over predicts the number of species with small abundance. This could be related to sampling, as the traps are less likely to capture multiple individuals of rare species, but it could also be related to the METE prediction for the number of singletons. METE predicts a number of singletons equal to βN_0, and therefore increasing N_0 while holding S_0 constant decreases the expected number of singletons (Harte 2011, Chapter 7.3). From Table 2, we see that N_0/S_0 is large for the pasture sites compared to the forest

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

sites, and therefore METE predicts proportionally fewer singletons at these sites. If ecologically we still expect a similar number of singletons, given that the high N_0/S_0 is being driven by a small number of very abundant species, then this could mean that METE under predicts the number of singletons.

4.2 Metabolic rate distribution of individuals (MRDI)

Across land use types, we find that the MRDIs are not particularly well described by the METE prediction. The Data section in Xiao et al. (2015) discusses why we should not necessarily expect animals (rather than plants) to follow the METE predicted MRDI. Namely, body sizes of animals belonging to the same species cluster around an intermediate value, and much larger or smaller individuals are rare (eg. Gouws et al. 2011). This means we are likely to end up with multimodal MRDIs (Thibault et al. 2011) rather than the monotonically decreasing form predicted by METE, which always predicts that the smallest size class is the most abundant. If we plot our metabolic rate distributions as histograms rather than by rank ordering, we do indeed find multimodality at most sites.

A further reason that we may not necessarily expect these distributions to fit well is the number of approximations required to obtain these distributions. We use scaling relationships to covert from body length to body mass (eg. Hódar 1996), and then approximate the intraspecific body mass distributions as normal, and then further use a scaling relationship to covert to metabolic rate. In addition, the body length-mass scaling relationships are used at the level of class, order, sub-order, or family (see Appendix S3), which masks likely significant variation within these taxonomic levels.

Yet another reason we may expect the empirical MRDIs to differ from the METE predictions here is that only adults are included in the metabolic rate distribution. This could result in missing the lower end of the unscaled MRDI, which will result in a skewed distribution. As METE predicts relative metabolic rate (scaled so that the smallest organism is $\varepsilon = 1$ (Harte 2011)), this will result in over predicting the metabolic rate of the individuals with the greatest metabolic rate. We see this in Fig. 3, in that at low rank the METE predictions are consistently too high.

Despite these concerns and the number of approximations, we can still quantify which land use is the most well described by METE, and we see a similar relationship between land use and goodness of fit when compared with the SAD and SAR.

At the pasture sites, we consistently under predict the low rank, high metabolic rate individuals, and over predict the high rank, low metabolic rate individuals, resulting in a pattern where the residuals have positive slope (Fig. 3). These patterns are indicative of a large number of individuals with similar metabolic rate (see Appendix S7 for plots of each site). As discussed in relation to

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the SAD, these sites have a few highly abundant, small bodied spider species. These species have comparatively low metabolic rate, and the variation in metabolic rate within a species is smaller than the variation across species. We therefore end up with long lines of positive slope in the residuals as the METE prediction slopes downward over rank but the empirical MRDI remains roughly constant. We see this especially at intermediate and low ranks as these species have low metabolic rates. Thus, this pattern is also likely driven by a few highly abundant species.

4.3 Species–area relationship (SAR)

The mean least squares comparisons for the SAR in Fig. 1 and Appendix S5: Fig. S1 are noticeably different from those for the SAD and MRDI. Again, here we find that the semi-natural pasture is the worst fit by METE, but it is not as dramatic as in the other cases and the fit is much closer to both forest sites. Additionally, we find that the intensive pasture is the best fit by METE, though it is comparable to other land use types given the error. However, the mean least squares is not the only goodness of fit metric. Particularly in the case of the pasture sites here, we see clear patterns in the residuals in Fig. 4 and Appendix S5: Fig. S2 that METE under predicts z and correspondingly over predicts the number of species at small scales. This is in line with our analysis of the SAD at the pasture sites, in that these sites have more high abundance species compared to the METE prediction (ie. the SAD is more hollow than the METE predicted log series). Overall, even though the mean least squares metric is smaller at the intensive pasture sites, the direction of the difference is more biased.

In general, the pasture sites correspond to larger N_0/S_0 than the forest sites (see Fig. 4 and Table 2). When using $D = \log(N_0/S_0)$ as a scale variable, this means that the pastures are testing a different scale compared to the forest sites. We see this in the residuals in Fig. 5, where the METE prediction for z is noticeably lower than the data points starting around $\log(N_0/S_0) \approx 2$, which is also where most of the pasture data points are clustered. This could indicate that the failure of METE to accurately predict the SAR is coming more from the underlying prediction of a log series abundance distribution, rather than from the species-level spatial abundance distribution prediction, as the log series prediction for the pasture sites under predicts the abundance of the most abundant species.

4.4 Indigenous and exotic species

In comparing the goodness of fit results in Fig. 1 and Fig. 6 across land use types and patterns, we find that the overall fit is largely comparable when indigenous and exotic species are considered together and separately, and for many of the patterns and land use types there is little difference

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

between the mean least squares for the indigenous and introduced species. This is in line with previous studies that found that exotic species were integrated with indigenous species in the Azorean arthropod communities by analyzing the interspecific abundance-occupancy relationship (Gaston et al. 2006; Rigal et al. 2013). In terms of METE, the maximum entropy inference technique should apply to any collection of entities, however they are categorized. For example, using arthropod data from Panama, Harte and Kitzes (2015) found evidence that the analysis of the form of the SAD is relatively insensitive to the choice or taxonomic category used for the analysis. Our findings here provide further support for a flexible application of METE across colonization categories.

For the differences that do exist, a possible interpretation is that the indigenous and introduced species fill different niches. In this case, they should be considered together in the context of METE as they form a single community to fill niche space. This could be the case because novel or unfilled niche space may make it easier for exotic species to invade, or for existing species to adapt to fill that space and expand their population size. Both of these examples would cause changes in the state variables when indigenous or introduced species are considered separately, and the dynamics may therefore move away from steady state. This would result in a worse fit for indigenous or introduced species when considered separate compared to together. Supporting this hypothesis, Rigal et al. (2018) found different functional trait profiles for indigenous and introduced species across the four different land uses considered, suggesting that introduced species are filling vacant ecological niches not occupied by indigenous species.

For the SADs, the results support our hypothesis that the poor fit at the semi-natural pasture sites is driven by introduced species, since outside of that site type the goodness of fit is comparable between indigenous and introduced species. That there is such a difference in goodness of fit for the SAD at the semi-natural pasture could therefore be pointing to more complex dynamics between the indigenous and introduced species at these sites (discussed in more depth below).

For the MRDI, that the semi-natural pasture is the best fit land use type when separating species is surprising given the results in Fig. 1. The relatively poor goodness of fit of the introduced species at the forest sites when compared to analyzing the species together may indicate that it is more important for the species to be analyzed together there, perhaps because these species are filling different niches as discussed above. Moreover, previous work has shown that the introduced species have a high spatial and temporal dynamism (ie. source sink dynamics) in the native forest, and their abundance and composition there are driven more by stochastic than equilibrium-type processes (Matthews et al. 2019; Borges et al. 2020). However, given that the potential error in calculating the MRDI is large, any result we see only in the MRDI should be interpreted cautiously.

For the SAR, it is interesting that we see the trend we initially expected, in that goodness of fit declines with increasing land use intensity for the indigenous species and increases with increasing land use intensity for the introduced species. However, the error bars are relatively large here, so

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

more data would be needed to interpret this result further.

Finally, an additional trend worth noting is that the goodness of fit for the indigenous species is better than for the introduced species at the native forest for both the MRDI and the SAR, while at the intensive pasture this is reversed and the introduced species fit better. This may be indicating that in these communities, the species more adapted for the habitat type tend to fit the METE predictions better. For example, in the native forest, as noted above, many of the human adapted introduced species are likely only present in a stochastic sense due to source-sink dynamics (Matthews et al. 2019), and thus may not be complete communities, leading to poor fit with METE predictions. However, this also may be due to smaller sample sizes at these sites as there are fewer of these species present.

4.5 Implications for future METE studies in anthropogenic landscapes

In other studies of METE, disturbance is often linked to rapid change in state variables (Newman et al. 2020; Franzman et al. 2021; Harte et al. 2021). The dynamics are then out of steady state, and the state variables alone are not sufficient to describe the macroecological patterns. For example, Franzman et al. (2021) analyzed the change in state variables over time in a declining alpine meadow and found that macroecological patterns moved away from METE predictions over a six year period of observation. Here, we instead analyze how land use change affects deviation from METE predictions assuming that the deviation in time at any given land use type is relatively static. Assuming that disturbance is connected to the rate of change of the state variables, we could interpret the poor fit of the semi-natural pasture as indicating that N_0, E_0 and/or S_0 are not constant on ecological time scales. We could test this hypothesis with time resolved data of arthropod composition. For example, we might expect the state variables to change with management intensity over the year. It also may be the case that disturbance is more general and cannot always be characterized by changing state variables, and may depend on additional factors such as the rate of migration in and out of the ecosystem rather than just the net difference.

More broadly, we hypothesize that heterogeneity in species composition may lead to deviations with METE. It is possible that when species from different groups interact, their dynamics are more complicated and violate the underlying METE assumption that N_0, E_0, and S_0 alone are adequate to characterize the larger scale patterns. These more complicated dynamics could, for example, be related to the mix of indigenous and introduced species, core and occasional species, or to source-sink dynamics (see Matthews 2021), depending on the habitat.

In this study, the generally poor fit at the semi-natural pasture site could result from these complex dynamics, particularly in comparison to the intensive pasture sites where species adjustment may be more extreme and composition more homogeneous. The semi-natural pasture sites

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

combine more similar numbers of species from their native habitats and human adapted (largely introduced) species when compared to the intensive pasture sites, which are more weighted in terms of the latter (Table 2). The semi-natural pasture sites are also in greater proximity to forest sites, which could impact how species use these sites in terms of complex source-sink dynamics (Borges and Brown 2004; Matthews 2021). For example, perhaps some species, particularly indigenous species, use the semi-natural pasture to disperse between different fragmented forest habitats (Borges et al. 2008).

Overall then, semi-natural pasture sites may combine many different types of abundant species; types which have difference dynamics and thus leave different imprints on the emergent macroecological patterns (Matthews 2021). The effect of combining different types or groups of species (eg. core-occasional, indigenous-introduced, larger taxonomic breadth) into a single sample on SAD form is well known (eg. Magurran and Henderson 2003; Matthews and Whittaker 2015; Antão et al. 2017), though their effect in the context of METE remains to be explored further.

4.6 Summary

Across METE predictions, and when considering all species together, the forest habitats are better predicted by METE than the semi-natural pasture habitat. The intensive pasture is intermediately well fit for the SAD and MRDI, and better fit for the SAR, though the residuals are not normally distributed.

For the forest sites, the native forest has no human management but may be subject to the spread of invasive plants, and the exotic forest is subject to some human management and is in close proximity to the pastures. The deviations from METE in these sites are comparatively small and there are less noticeable trends in the residuals, though as with all sites METE over predicts the metabolic rate of the highest metabolic rate individuals.

The pasture sites are characterized by a few very abundant species, which is consistent with the abundance of several small bodied introduced spiders. In the SAD, METE consistently under predicts the abundance of the most abundant species, as well as the number of singletons. In the MRDI, these very abundant species appear in the residuals as long lines of positive slope, as the variation in metabolic rate within a species is relatively small and the METE prediction falls off more rapidly than the empirical distribution.

The semi-natural pasture is particularly poorly described by METE across metrics. This is also the only land use with a large difference in goodness of fit between indigenous and introduced species (when analyzed separately) for the SAD, and the poor fit for the SAD appears to be driven by introduced species. This could be due to high source-sink dynamics and complex interactions between indigenous and introduced species, particularly because of the proximity to other land

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

uses, or because of the varying levels of management in semi-natural pastures over the course of a year.

The comparatively better fit at the intensive pasture site could result from the sensitive species having already been lost and thus the remaining arthropod communities comprising species that are better adapted to the high level of management intensity. In terms of METE, this may mean that state variables change less rapidly at these sites, or perhaps the more homogeneous species composition means interactions are simpler despite the higher degree of disturbance.

Analyzing the deviation from METE predictions across land use has provided us with useful information about how land use and related disturbance is affecting macroecological patterns in Azorean arthropods. While we initially expected the intensive pasture sites to be the most poorly fit by METE, this analysis points to the semi-natural pasture as the land use where arthropod communities are the most out of steady state. We were additionally able to interpret the deviations from METE predictions ecologically. We expect this type of comparison between METE predictions and ecosystems under land management disturbance to be helpful in identifying how land use affects macroecological patterns across other habitats and taxa.

Acknowledgments

We would like to thank Rosemary Gillespie for her insights and helpful feedback. Funding for this project was provided in part by grant DEB 1751380 from the US National Science Foundation, as well as by grant FCT-UIDP/00329/2020-2024 (Thematic Line 1–Integrated ecological assessment of environmental change on biodiversity) through the FCT – Fundação para a Ciência e a Tecnologia. MB acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC) [PGSD2-517114-2018]. Data acquisition was provided by the projects: "Consequences of land use change on Azorean fauna and flora – the 2010 Target" (Ref: Direcção Regional de Ciência e Tecnologia M.2.1.2/I/003/2008) and 'Direcção Regional dos Recursos Florestais' ('Secretaria Regional da Agricultura e Pescas') through the Project 'Reservas Florestais dos Açores: Cartografia e Inventariação dos Artrópodes Endémicos dos Açores' (PROJ. 17.01 – 080203).

Literature Cited

Antão, L. H., S. R. Connolly, A. E. Magurran, A. Soares, and M. Dornelas (2017). "Prevalence of multimodal species abundance distributions is linked to spatial and taxonomic breadth". *Global Ecology and Biogeography* 26.2, pp. 203–215.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Arnold, T. B. and J. W. Emerson (2011). "Nonparametric Goodness-of-Fit Tests for Discrete Null Distributions". *The R Journal* 3.2, pp. 34–39.

Baumgärtner, D. and K.-O. Rothhaupt (2003). "Predictive Length–Dry Mass Regressions for Freshwater Invertebrates in a Pre-Alpine Lake Littoral". *International Review of Hydrobiology* 88.5, pp. 453–463.

Borda-de-Água, L. et al. (2017). "Dispersal ability determines the scaling properties of species abundance distributions: a case study using arthropods from the Azores". *Scientific Reports* 7.1, p. 3899.

Borges, P. A. V. and V. K. Brown (2004). "Arthropod community structure in pastures of an island archipelago (Azores): looking for local-regional species richness patterns at fine-scales". *Bulletin of Entomological Research* 94.2, pp. 111–121.

Borges, P. A. V., J. M. Lobo, E. B. d. Azevedo, C. S. Gaspar, C. Melo, and L. V. Nunes (2006). "Invasibility and species richness of island endemic arthropods: a general model of endemic vs. exotic species". *Journal of Biogeography* 33.1, pp. 169–187.

Borges, P. A. V., F. Rigal, A. Ros-Prieto, and P. Cardoso (2020). "Increase of insular exotic arthropod diversity is a fundamental dimension of the current biodiversity crisis". *Insect Conservation and Diversity* 13.5, pp. 508–518.

Borges, P. A. V., K. I. Ugland, F. Dinis, and C. Gaspar (2008). "Insect and spider rarity in an oceanic island (Terceira, Azores): true rare and pseudo-rare species". *Insect Ecology and Conservation*. Ed. by S. Fattorini. Kerala (India): Research Signpost, pp. 47–70.

Borges, P. A. V. and J. Wunderlich (2008). "Spider biodiversity patterns and their conservation in the Azorean archipelago, with descriptions of new species". *Systematics and Biodiversity* 6.2, pp. 249–282.

Borges, P. A. V. et al. (2005). "Ranking protected areas in the Azores using standardised sampling of soil epigean arthropods". *Biodiversity & Conservation* 14.9, pp. 2029–2060.

Borges, P. A. V. et al. (2010). "List of arthropods (Arthropoda)". *A list of the terrestrial and marine biota from the Azores*. Ed. by P. A. V. Borges et al. Parede: Principia, pp. 179–246.

Brummer, A. B. and E. A. Newman (2019). "Derivations of the Core Functions of the Maximum Entropy Theory of Ecology". *Entropy* 21.7, p. 712.

Cardoso, P., S. C. Aranda, J. M. Lobo, F. Dinis, C. Gaspar, and P. A. V. Borges (2009). "A spatial scale assessment of habitat effects on arthropod communities of an oceanic island". *Acta Oecologica* 35.5, pp. 590–597.

Carey, S., J. Harte, and R. D. Moral (2006). "Effect of community assembly and primary succession on the species-area relationship in disturbed ecosystems". *Ecography* 29.6, pp. 866–872.

Conlisk, E., M. Bloxham, J. Conlisk, B. Enquist, and J. Harte (2007). "A New Class of Models of Spatial Distribution". *Ecological Monographs* 77.2, pp. 269–284.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Connolly, S. R. and M. Dornelas (2011). "Fitting and empirical evaluation of models for species abundance distributions". *Biological Diversity: Frontiers in Measurement and Assessment*. Ed. by A. E. Magurran and B. J. McGill. Oxford, UNITED KINGDOM: Oxford University Press.

Dornelas, M., A. C. Moonen, A. E. Magurran, and P. Bàrberi (2009). "Species abundance distributions reveal environmental heterogeneity in modified landscapes". *Journal of Applied Ecology* 46.3, pp. 666–672.

Elias, R. B., A. Gil, L. Silva, J. M. Fernández-Palacios, E. B. Azevedo, and F. Reis (2016). "Natural zonal vegetation of the Azores Islands: characterization and potential distribution". *Phytocoenologia* 46.2, pp. 107–123.

Fahrig, L. (2003). "Effects of Habitat Fragmentation on Biodiversity". *Annual Review of Ecology, Evolution, and Systematics* 34.1, pp. 487–515.

— (2019). "Habitat fragmentation: A long and tangled tale". *Global Ecology and Biogeography* 28.1, pp. 33–41.

Fattorini, S., F. Rigal, P. Cardoso, and P. A. V. Borges (2016). "Using species abundance distribution models and diversity indices for biogeographical analyses". *Acta Oecologica* 70, pp. 21–28.

Fischer, J. and D. B. Lindenmayer (2007). "Landscape modification and habitat fragmentation: a synthesis". *Global Ecology and Biogeography* 16.3, pp. 265–280.

Florencio, M., P. Cardoso, J. M. Lobo, E. B. d. Azevedo, and P. A. V. Borges (2013). "Arthropod assemblage homogenization in oceanic islands: the role of indigenous and exotic species under landscape disturbance". *Diversity and Distributions* 19.11, pp. 1450–1460.

Foley, J. A. et al. (2005). "Global Consequences of Land Use". *Science* 309.5734, pp. 570–574.

Franzman, J., M. Brush, K. Umemura, C. Ray, B. Blonder, and J. Harte (2021). "Shifting macroecological patterns and static theory failure in a stressed alpine plant community". *Ecosphere* 12.6.

Gaspar, C., P. A. V. Borges, and K. J. Gaston (2008). "Diversity and distribution of arthropods in native forests of the Azores archipelago". *Arquipélago* Life and Marine Sciences.25, p. 30.

Gaston, K. J., P. A. V. Borges, F. He, and C. Gaspar (2006). "Abundance, spatial variance and occupancy: arthropod species distribution in the Azores". *Journal of Animal Ecology* 75.3, pp. 646–656.

Gillespie, R. G., E. M. Claridge, and G. K. Roderick (2008). "Biodiversity dynamics in isolated island communities: interaction between natural and human-mediated processes". *Molecular Ecology* 17.1, pp. 45–57.

Gillespie, R. G. and G. K. Roderick (2002). "Arthropods on Islands: Colonization, Speciation, and Conservation". *Annual Review of Entomology* 47.1, pp. 595–632.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gouws, E. J., K. J. Gaston, and S. L. Chown (2011). "Intraspecific Body Size Frequency Distributions of Insects". *PLOS ONE* 6.3, e16606.

Gray, J. S. et al. (1979). "Pollution-induced changes in populations". *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 286.1015, pp. 545–561.

Gray, J. S., A. Bjørgesæter, and K. I. Ugland (2006). "On plotting species abundance distributions". *Journal of Animal Ecology* 75.3, pp. 752–756.

Harte, J., T. Zillio, E. Conlisk, and A. B. Smith (2008). "Maximum Entropy and the State-Variable Approach to Macroecology". *Ecology* 89.10, pp. 2700–2711.

Harte, J. (2011). *Maximum entropy and ecology : a theory of abundance, distribution, and energetics*. Oxford Series in Ecology and Evolution. Oxford: Oxford University Press. 257 pp.

Harte, J. and J. Kitzes (2015). "Inferring Regional-Scale Species Diversity from Small-Plot Censuses". *PLOS ONE* 10.2, e0117527–e0117527.

Harte, J. and E. A. Newman (2014). "Maximum information entropy: a foundation for ecological theory". *Trends in Ecology & Evolution* 29.7, pp. 384–389.

Harte, J., A. B. Smith, and D. Storch (2009). "Biodiversity scales from plots to biomes with a universal species-area curve". *Ecology Letters* 12, pp. 789–797.

Harte, J., K. Umemura, and M. Brush (2021). "DynaMETE: a hybrid MaxEnt-plus-mechanism theory of dynamic macroecology". *Ecology Letters* 24.5, pp. 935–949.

Hill, J. K. and K. C. Hamer (1998). "Using species abundance models as indicators of habitat disturbance in tropical forests". *Journal of Applied Ecology* 35.3, pp. 458–460.

Hódar, J. A. (1996). "The use of regression equations for estimation of arthropod biomass in ecological studies". *Acta Oecologica* 17.5, pp. 421–433.

Hubbell, S. P. (2013). "Tropical rain forest conservation and the twin challenges of diversity and rarity". *Ecology and Evolution* 3.10, pp. 3263–3274.

Kempton, R. A. and L. R. Taylor (1974). "Log-Series and Log-Normal Parameters as Diversity Discriminants for the Lepidoptera". *Journal of Animal Ecology* 43.2, pp. 381–399.

Kitzes, J. and M. Wilber (2016). "macroeco: reproducible ecological pattern analysis in Python". *Ecography* 39.4, pp. 361–367.

Kitzes, J., M. Wilber, C. Lewis, and E. P. White (2015). *jkitzes/macroeco*. Version 1.01.

Klein Goldewijk, K., A. Beusen, J. Doelman, and E. Stehfest (2017). "Anthropogenic land use estimates for the Holocene – HYDE 3.2". *Earth System Science Data* 9.2, pp. 927–953.

Macías-Hernández, N. et al. (2020). "A database of functional traits for spiders from native forests of the Iberian Peninsula and Macaronesia". *Biodiversity Data Journal* 8, e49159.

Magurran, A. E. and P. A. Henderson (2003). "Explaining the excess of rare species in natural species abundance distributions". *Nature* 422.6933, pp. 714–716.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Martins, I. S., V. Proença, and H. M. Pereira (2014). "The unusual suspect: Land use is a key predictor of biodiversity patterns in the Iberian Peninsula". *Acta Oecologica* 61, pp. 41–50.

Matthews, T. J. and R. J. Whittaker (2014). "Fitting and comparing competing models of the species abundance distribution: assessment and prospect". *Frontiers of Biogeography* 6.2.

Matthews, T. J. (2021). "On The Biogeography of Habitat Islands: The Importance of Matrix Effects, Noncore Species, and Source-Sink Dynamics". *The Quarterly Review of Biology* 96.2, pp. 73–104.

Matthews, T. J., J. Sadler, R. Carvalho, R. Nunes, and P. A. V. Borges (2019). "Differential temporal beta-diversity patterns of native and non-native arthropod species in a fragmented native forest landscape". *Ecography* 42.1, pp. 45–54.

Matthews, T. J. and R. J. Whittaker (2015). "On the species abundance distribution in applied ecology and biodiversity management". *Journal of Applied Ecology* 52.2, pp. 443–454.

Maxwell, S. L., R. A. Fuller, T. M. Brooks, and J. E. M. Watson (2016). "Biodiversity: The ravages of guns, nets and bulldozers". *Nature* 536.7615, pp. 143–145.

McGill, B. J. et al. (2007). "Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework". *Ecology Letters* 10.10, pp. 995–1015.

McGill, B. J. et al. (2019). "Unifying macroecology and macroevolution to answer fundamental questions about biodiversity". *Global Ecology and Biogeography* 28.12, pp. 1925–1936.

Newbold, T. et al. (2015). "Global effects of land use on local terrestrial biodiversity". *Nature* 520.7545, pp. 45–50.

Newbold, T. et al. (2018). "Widespread winners and narrow-ranged losers: Land use homogenizes biodiversity in local assemblages worldwide". *PLOS Biology* 16.12, e2006841.

Newman, E. A. (2019). "Disturbance Ecology in the Anthropocene". *Frontiers in Ecology and Evolution* 7.

Newman, E. A. et al. (2020). "Disturbance macroecology: a comparative study of community structure metrics in a high-severity disturbance regime". *Ecosphere* 11.1, e03022.

Norder, S. J. et al. (2020). "Global change in microcosms: Environmental and societal predictors of land cover change on the Atlantic Ocean Islands". *Anthropocene* 30, p. 100242.

Pereira, H. M., L. M. Navarro, and I. S. Martins (2012). "Global Biodiversity Change: The Bad, the Good, and the Unknown". *Annual Review of Environment and Resources* 37.1, pp. 25–50.

Pimm, S. L. et al. (2014). "The biodiversity of species and their rates of extinction, distribution, and protection". *Science* 344.6187.

Rigal, F., R. J. Whittaker, K. A. Triantis, and P. A. V. Borges (2013). "Integration of non-indigenous species within the interspecific abundance–occupancy relationship". *Acta Oecologica* 48, pp. 69–75.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Rigal, F. et al. (2018). "Functional traits of indigenous and exotic ground-dwelling arthropods show contrasting responses to land-use change in an oceanic island, Terceira, Azores". *Diversity and Distributions* 24.1, pp. 36–47.

Rominger, A. J. et al. (2016). "Community assembly on isolated islands: macroecology meets evolution". *Global Ecology and Biogeography* 25.7, pp. 769–780.

Simons, N. K., M. M. Gossner, T. M. Lewinsohn, M. Lange, M. Türke, and W. W. Weisser (2015). "Effects of land-use intensity on arthropod species abundance distributions in grasslands". *Journal of Animal Ecology* 84.1, pp. 143–154.

Steege, H. ter et al. (2013). "Hyperdominance in the Amazonian Tree Flora". *Science* 342.6156, pp. 1243092–1243092.

Supp, S. R., X. Xiao, S. K. M. Ernest, and E. P. White (2012). "An experimental test of the response of macroecological patterns to altered species interactions". *Ecology* 93.12, pp. 2505–2511.

Terzopoulou, S., F. Rigal, R. J. Whittaker, P. A. V. Borges, and K. A. Triantis (2015). "Drivers of extinction: the case of Azorean beetles". *Biology Letters* 11.6, p. 20150273.

Thibault, K. M., E. P. White, A. H. Hurlbert, and S. K. M. Ernest (2011). "Multimodality in the individual size distributions of bird communities". *Global Ecology and Biogeography* 20.1, pp. 145–153.

Ulrich, W., M. Ollik, and K. I. Ugland (2010). "A meta-analysis of species–abundance distributions". *Oikos* 119.7, pp. 1149–1155.

Wardhaugh, C. W. (2013). "Estimation of biomass from body length and width for tropical rainforest canopy invertebrates". *Australian Journal of Entomology* 52.4, pp. 291–298.

White, E. P., K. M. Thibault, and X. Xiao (2012). "Characterizing species abundance distributions across taxa and ecosystems using a simple maximum entropy model". *Ecology* 93.8, pp. 1772–1778.

Whittaker, R. J. and J. M. Fernández-Palacios (2007). *Island biogeography: ecology, evolution, and conservation*. 2nd ed. Oxford biology. Oxford ; New York: Oxford University Press. 401 pp.

Whittaker, R. J., J. M. Fernández-Palacios, T. J. Matthews, M. K. Borregaard, and K. A. Triantis (2017). "Island biogeography: Taking the long view of nature's laboratories". *Science* 357.6354.

Wilber, M. Q., J. Kitzes, and J. Harte (2015). "Scale collapse and the emergence of the power law species–area relationship". *Global Ecology and Biogeography* 24.8, pp. 883–895.

Williamson, M. and K. J. Gaston (2005). "The lognormal distribution is not an appropriate null hypothesis for the species–abundance distribution". *Journal of Animal Ecology* 74.3, pp. 409–422.

Xiao, X., D. J. Mcglinn, and E. P. White (2015). "A Strong Test of the Maximum Entropy Theory of Ecology". *The American Naturalist* 185.3, E70–E80.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Xu, C., Z. Wang, Z. Li, L. Wang, and G. Han (2019). "Grazing intensity and climate factors shape species abundance distribution by influencing different components of plant communities in a desert steppe". *Ecological Research* 34.6, pp. 730–742.

Zillio, T. and F. He (2010). "Modeling spatial aggregation of finite populations". *Ecology* 91.12, pp. 3698–3706.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Macroecological pattern	Description	METE predicted form
Species abundance distribution (SAD)	The probability distribution of species with abundance n. Describes the commonness and rarity of species.	$\Phi(n) = \frac{e^{-\beta n}}{n \log\left(1/\left(1-e^{-\beta}\right)\right)}$
Metabolic rate distribution of individuals (MRDI)	The probability distribution of individuals with metabolic rate ε. Describes the metabolic make up of the community. Can be converted to body size using metabolic scaling.	$\Psi(\varepsilon) = \lambda_2 \left(e^{\beta} - 1\right) \frac{e^{-\gamma}}{(1-e^{-\gamma})^2}$
Species–area relationship (SAR)	Describes the relationship between the area of an ecosystem and the number of species found within that area.	$S(A) = \sum_{n_0} \Phi(n_0) \left(1 - \Pi(0\|A, A_0, n_0)\right)$

Table 1: Descriptions of the macroecological patterns used in this study. The forms predicted by the Maximum Entropy Theory of Ecology are shown in the right column, where λ_1 and λ_2 are the Lagrange multipliers, $\beta = \lambda_1 + \lambda_2$, and $\gamma = \lambda_1 + \lambda_2\varepsilon$. The species–area relationship must be calculated at each new scale A from an existing scale A_0 using the species abundance distribution $\Phi(n)$ and the species-level spatial abundance distribution $\Pi(n)$.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Land use	% Area	Sites	Total S_0	Total N_0	Median S_0	Median N_0
Native forest	9	44	148 (86, 60)	10 291 (7288, 3001)	24 (16, 8)	195 (129, 50)
Exotic forest	15	12	87 (44, 42)	3385 (1476, 1908)	20 (10, 9)	196 (11, 51)
Semi-natural pasture	15	16	127 (50, 76)	11 421 (2110, 9310)	28 (10, 17)	766 (101, 623)
Intensive pasture	48	24	136 (40, 94)	21 153 (4076, 17 070)	36 (10, 27)	878 (161, 684)

Table 2: The total number of species and individuals observed for each land use, and the median number across sites within one land use. The number in parentheses is the number of indigenous species or individuals, followed by the number of exotic species or individuals. Additionally, the number of sites where data were collected for each land use, and the percent of the total island area occupied by that land use. Across all land uses, there are a total of 271 species and 46 250 individuals, with four species constituting 11 individuals that are not identified as indigenous or exotic. The dataset in full is used for the SAD analysis, but for the MRDI analysis juvenile individuals were excluded, leaving a total of 226 species and 36 269 individuals, and for the SAR analysis there are less data available with the spatial resolution needed, leaving 228 species and 34 282 individuals.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

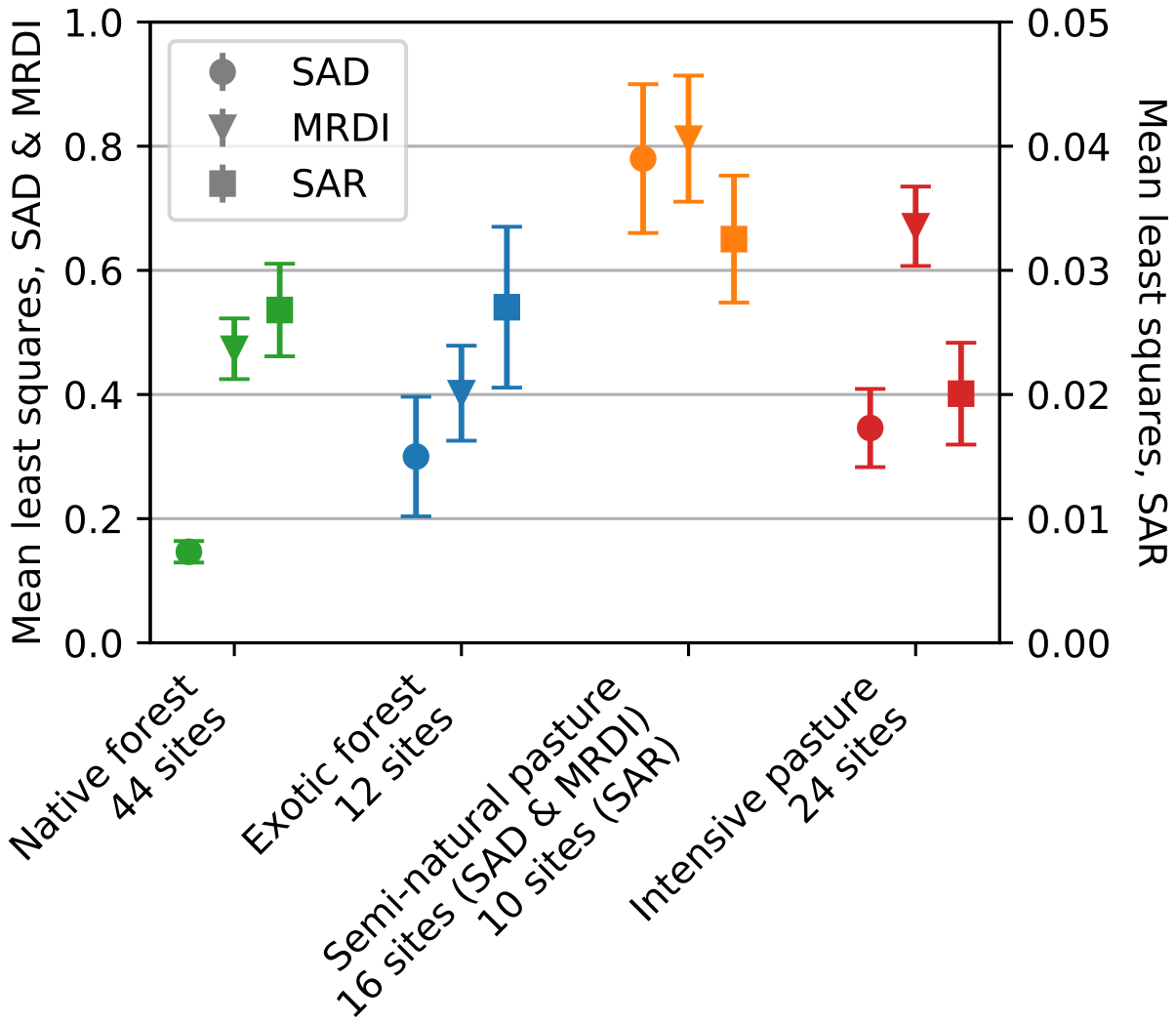

Figure 1: A comparison of means (averaged across the results from individuals transects) of the mean least squares and their standard errors across all three patterns (SAD, MRDI, SAR) and land use types. For the SAD and the MRDI, the mean least squared difference is the observed minus the predicted for the rank ordered natural log of the abundance or metabolic rate, respectively. For the SAR, the mean least squared difference is the observed minus the predicted for the slope, z. Note the difference in y-scale for the SAR, where the mean least squared error was much smaller. The shape of the marker indicates the pattern and the color indicates the land use.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

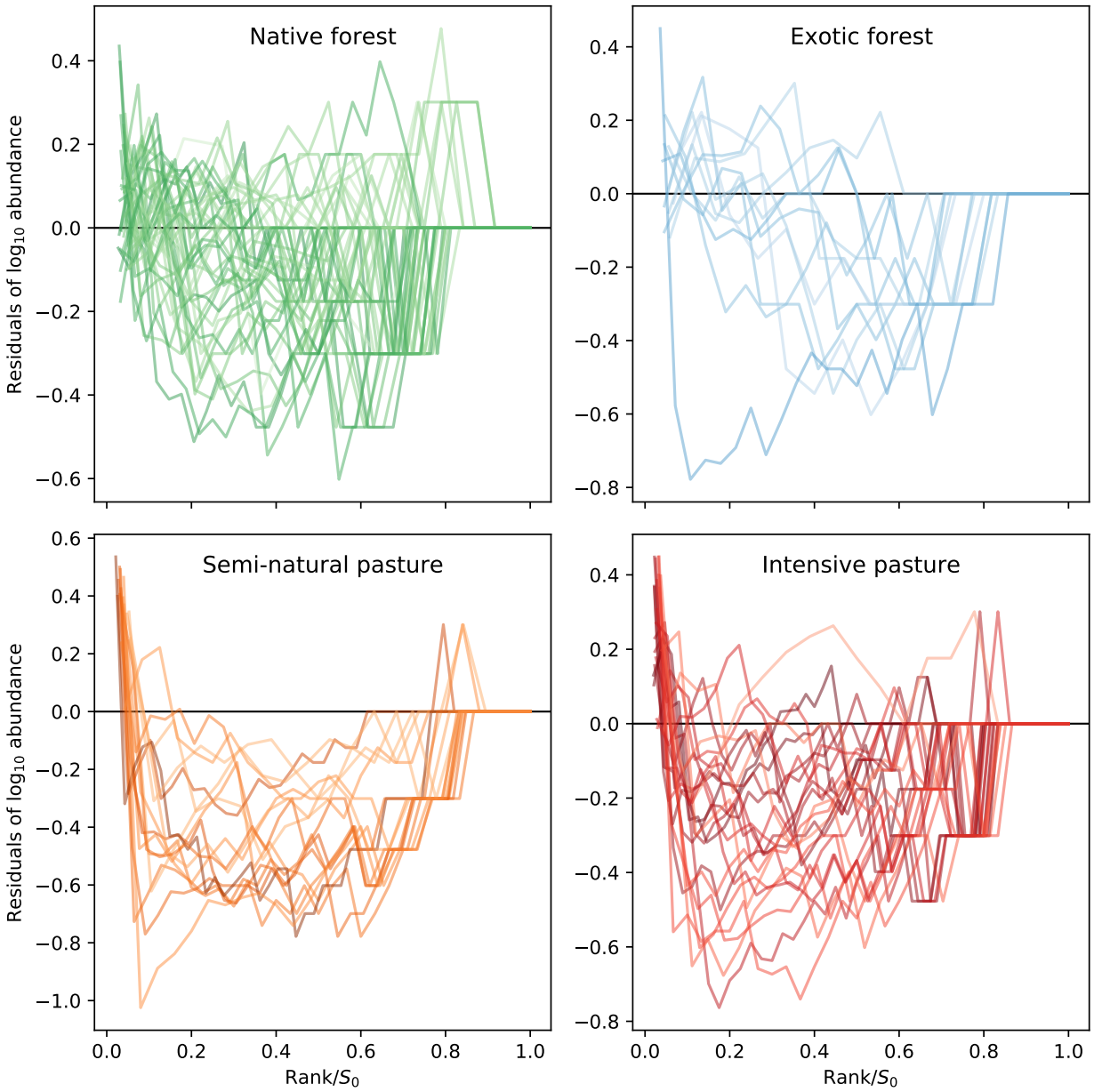

Figure 2: The observed $\log_{10}$ of abundance minus the predicted $\log_{10}$ of abundance from METE for each transect across land uses. As the number of ranks is equal to the number of species S_0, the ranks on the x-axis have been rescaled by $1/S_0$ to facilitate comparison between sites. The darker lines are sites with a higher number of species, and lighter lines represent sites with fewer species. The colors correspond to the different land uses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

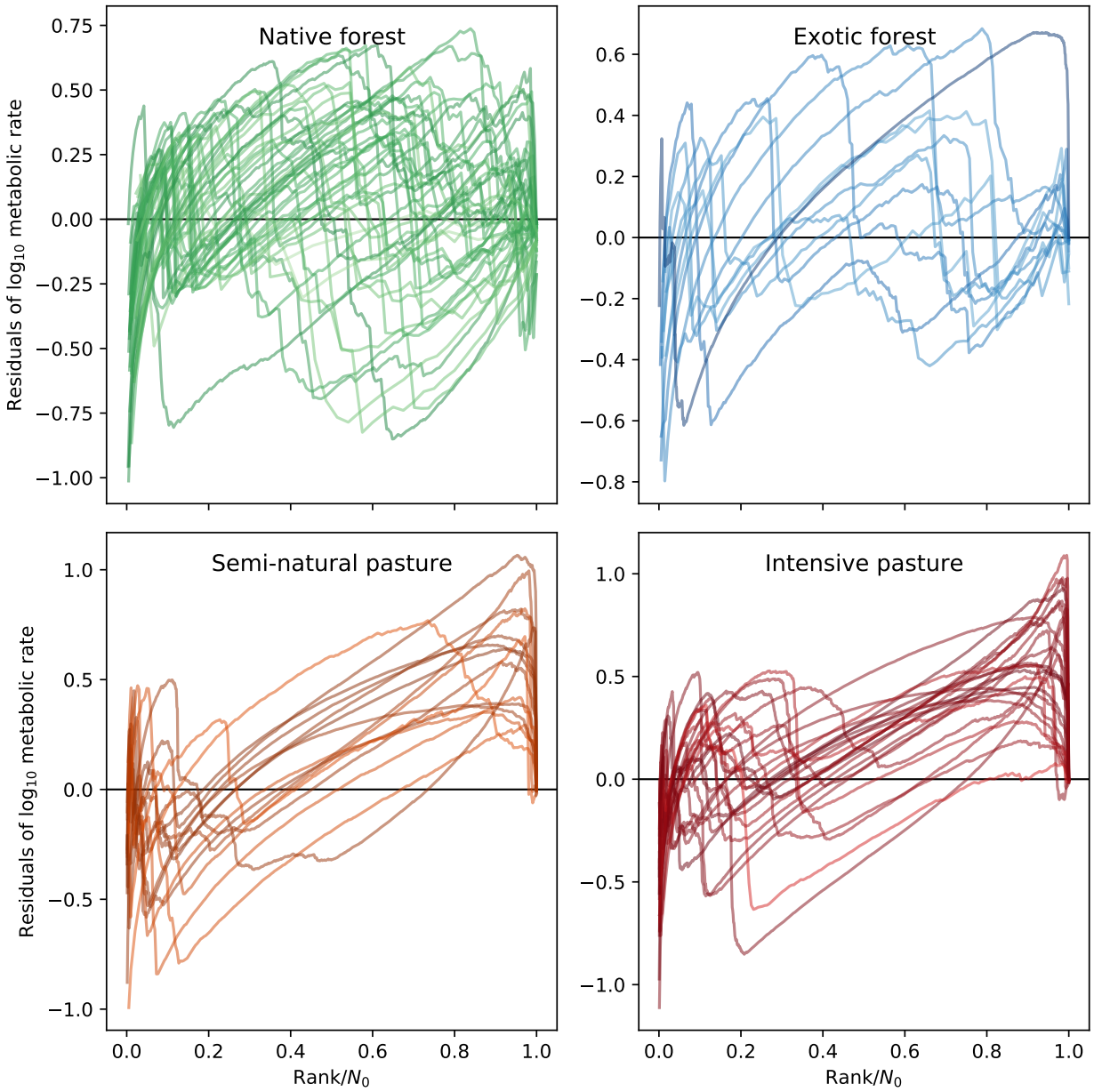

Figure 3: The observed $\log_{10}$ of the metabolic rate minus predicted $\log_{10}$ of the metabolic rate for the rank ordered plots. As the number of ranks is equal to the number of species S_0, the ranks on the x-axis have been rescaled by $1/S_0$ to facilitate comparison between sites. The darker lines are sites with a higher number of individuals, and lighter lines represent sites with fewer individuals. The colors correspond to the different land uses.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

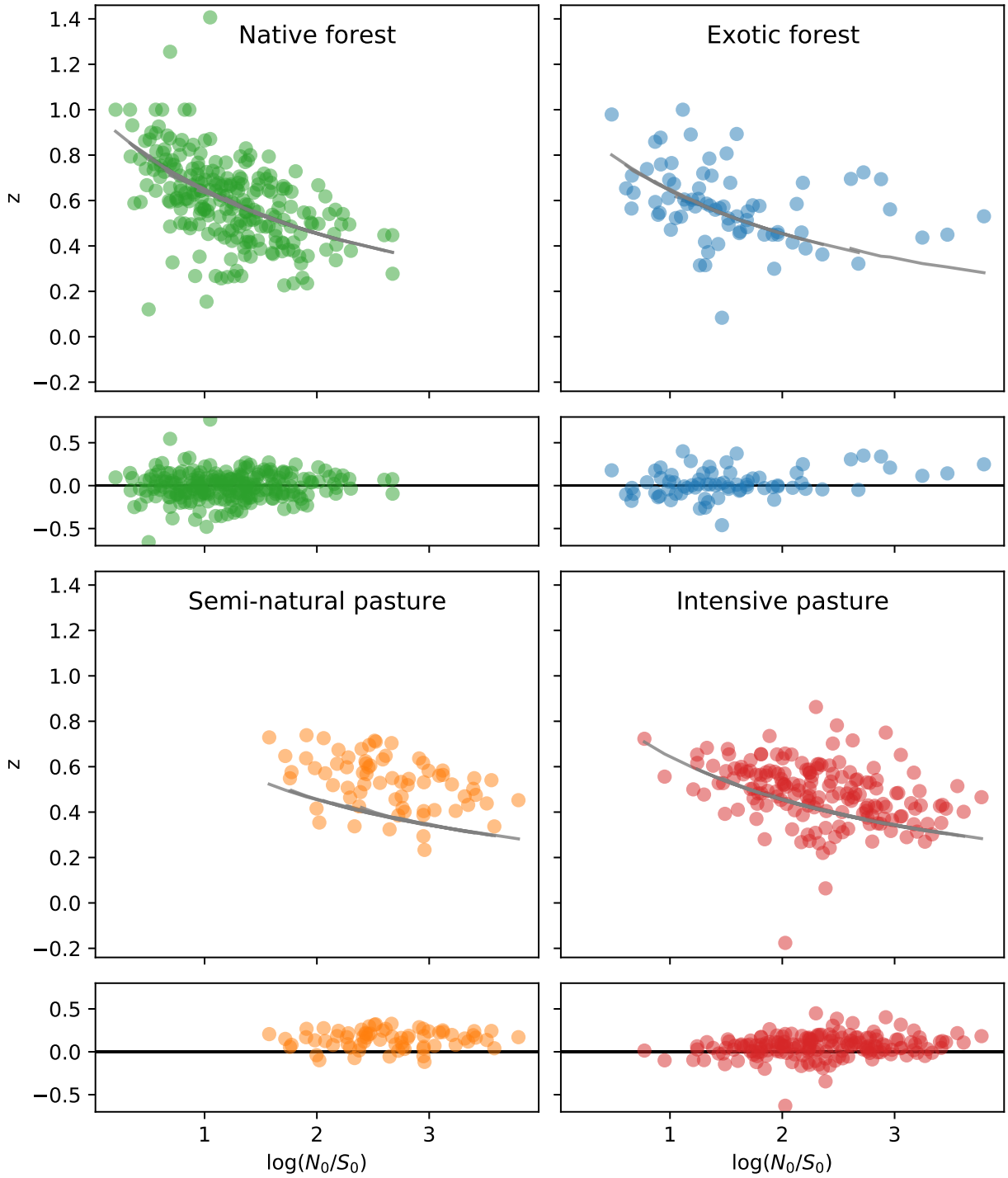

Figure 4: The species–area relationship for each transect across land use. Each point represents a single transect at a specific scale, where the scale is determined by $D = \log(N_0/S_0)$, and are colored according to land use type. The gray lines are the METE predictions, which largely overlap due to the scale collapse prediction of METE. Here we have plotted the slope of the relationship on the y-axis so that all points collapse onto one universal curve. The residuals ($z_{\text{observed}} - z_{\text{predicted}}$) for each land use are shown immediately below the plot for that land use.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

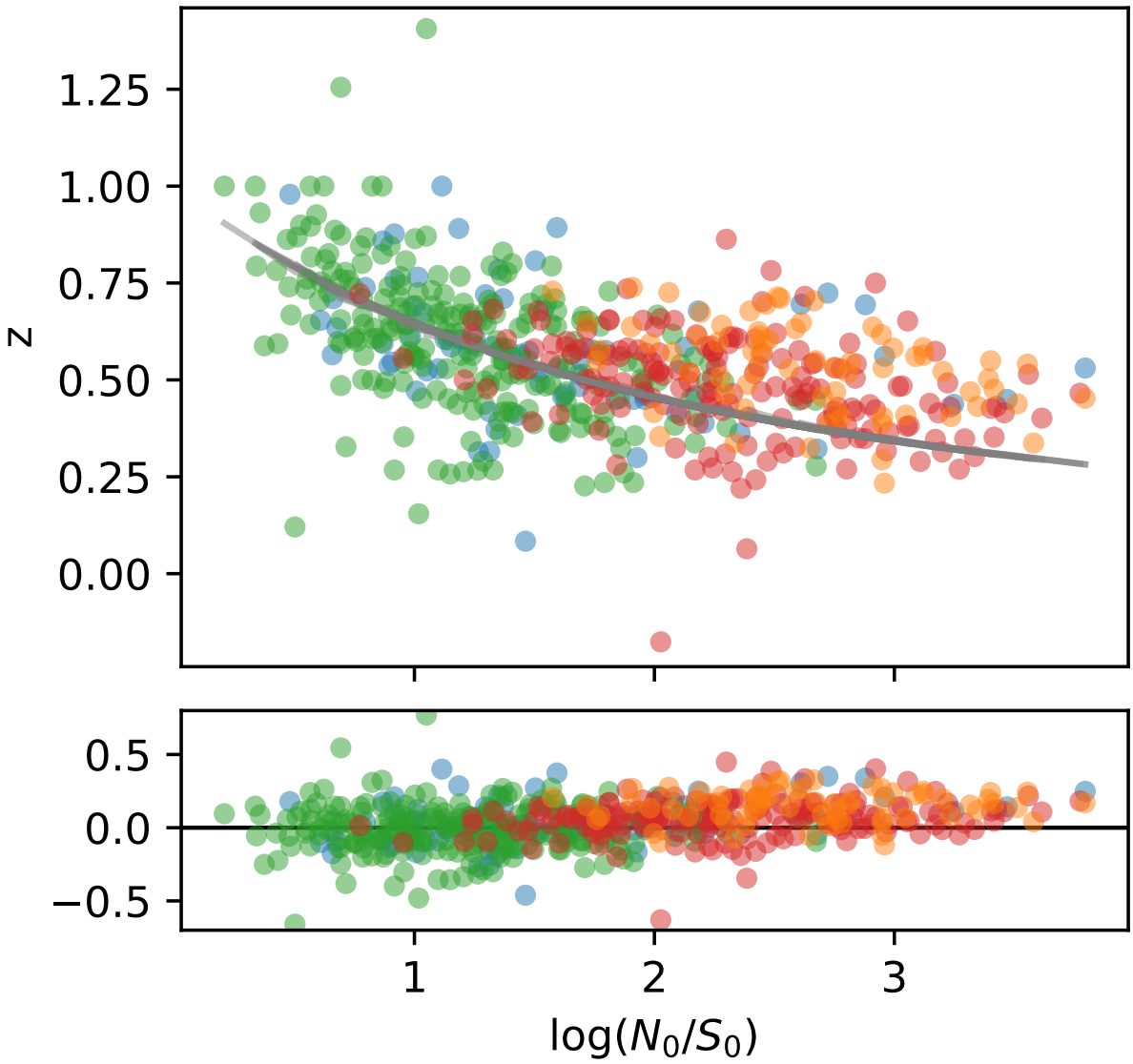

Figure 5: The species–area relationship for each transect all on a single plot, color coded by land use (colours match those in Fig. 4). Each point represents a single transect at a specific scale, where the scale is determined by $\log(N_0/S_0)$. The gray lines correspond to the METE predictions. Here we have plotted the slope of the relationship on the y-axis so that all points collapse onto one universal curve. The residuals are shown below.

bioRxiv preprint doi: https://doi.org/10.1101/2021.09.14.460355; this version posted September 17, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

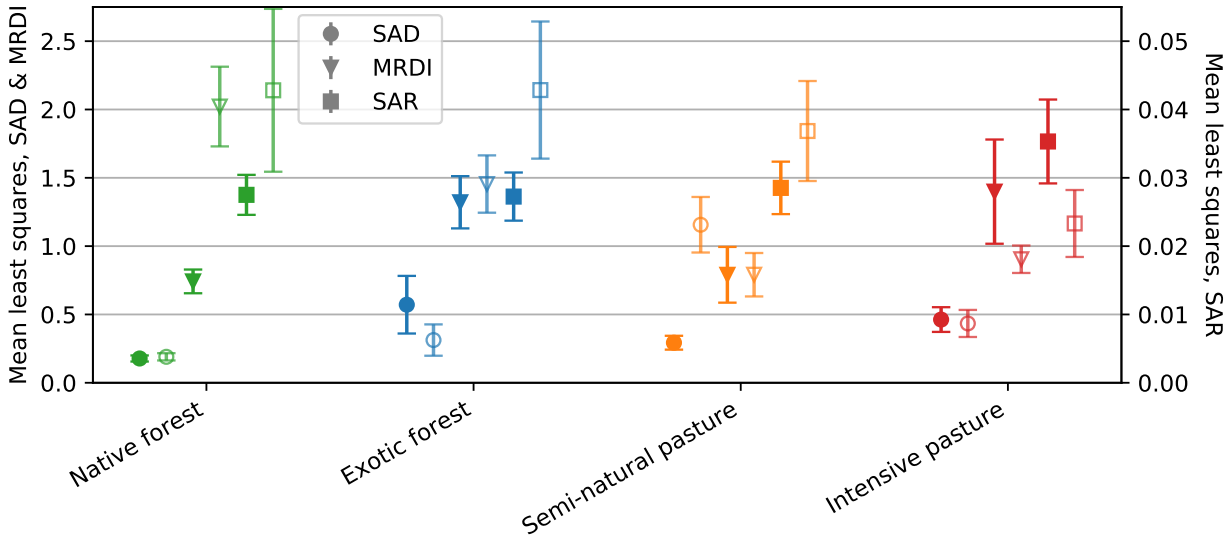

Figure 6: Mean and standard error of the mean of the mean least squares for all patterns across transects for each land use, for both indigenous (filled shapes) and introduced species (open shapes, lighter color). The definitions of mean least squared are as in Fig. 1, along with the meanings of the shapes and colors. The number of sites depends on the land use type, the pattern, and if the species are indigenous or introduced. For the SAD, the number of sites are as in Fig. 1, so 44, 12, 16, 24, in order of increasing land use intensity. For the MRDI, the number of sites are (44,32), (11,12), (15,16), (24,24), where the first number in parentheses is the number of sites with indigenous species and the second with introduced. For the SAR, the number of sites are (43,9), (10,9), (9,10),(20,24).