

## ORIGINAL ARTICLE

## Journal Section

# Derivations of the Core Functions of the Maximum Entropy Theory of Ecology

Alexander B. Brummer<sup>1</sup> | Erica A. Newman<sup>2\*</sup>

<sup>1</sup>Department of Biomathematics, David Geffen School of Medicine, University of California, Los Angeles

<sup>2</sup>Department of Ecology and Evolutionary Biology, University of Arizona

**Correspondence**

Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, 85721, USA

Email: newmane@email.arizona.edu

**Funding information**

Funding was provided in part by the USDA and the Bridging Biodiversity and Conservation Science program to EAN.

The Maximum Entropy Theory of Ecology, or METE, is a theoretical framework of macroecology that makes a variety of realistic ecological predictions about how species richness, abundance of species, metabolic rate distributions, and spatial aggregation of species interrelate in a given region. In the METE framework, “ecological state variables” (representing total area, total species richness, total abundance, and total metabolic energy) describe macroecological properties of an ecosystem. METE incorporates these state variables into constraints on underlying probability distributions. The method of Lagrange multipliers and maximization of information entropy (MaxEnt) lead to predicted functional forms of distributions of interest. We demonstrate how information entropy is maximized for the general case of a distribution, which has empirical information that provides constraints on the overall predictions. We then show how METE’s two core functions are derived. These functions, called the “Spatial Structure Function” and the “Ecosystem Structure Function” are the core pieces of the theory, from which all the predictions of METE follow (including the Species Area Distribution, the Species Abundance Distribution, and various metabolic distributions). Primarily, we consider the discrete distributions predicted by METE. We also explore the parameter space defined by the METE’s state variables and Lagrange multipliers. We aim to provide a comprehensive resource for ecologists who

want to understand the derivations and assumptions of basic mathematical structure of METE.

#### KEY WORDS

information entropy, information theoretics, macroecology, metabolic theory, scaling, species abundance distribution, species-area relationship

## 1 | THE MAXIMUM ENTROPY THEORY OF ECOLOGY

Many of the central questions of macroecology ask how patterns of species richness, abundance, and body size arise from ecosystems, how these patterns scale over increasing area, and how they interrelate [1]. Many macroecological distributions that quantify aspects of community structure, such as the Species-Area Relationship, the Species Abundance Distribution [2], size-density relationships [3, 4, 5], and the allometric scaling of metabolic rates of biological organisms within a community [6, 7, 8] have been studied independently, revealing general properties that may be universal across ecosystems. The Maximum Entropy Theory of Ecology, or METE [9, 10, 11], is a theoretical framework of macroecology that makes a variety of realistic ecological predictions about the diversity and structure of ecological communities [12, 13, 14, 15, 16, 17, 18, 19, 20]. These predictions relate species richness and abundance to metabolic rate distributions and spatial aggregation of species in a given region. Because METE makes a set of interrelated predictions about community structure, it has the potential to unify disparate parts of macroecology into a single mathematical framework.

The underlying mathematics of METE relies on a method termed "MaxEnt": the maximization of information entropy. MaxEnt uses the method of Lagrange multipliers to find probability distributions that underlie statistical phenomena. For METE, the MaxEnt method is applied to problems involving measurable "ecological state variables" that describe macroecological properties of an ecosystem.

In this paper, we will first demonstrate how information entropy is maximized for the general case of a distribution, which has empirical information that provides constraints on the overall predictions. We then introduce the ecological state variables  $A_0$ ,  $S_0$ ,  $N_0$ ,  $E_0$ , representing total area, total number of species, total abundance, and total metabolic energy of an ecological system, and use them with the method of information entropy maximization to show how METE's two core functions are derived. These functions, called the "Spatial Structure Function" and the "Ecosystem Structure Function" are the core pieces of the theory, from which all the predictions of METE follow (including the Species-Area Distribution, the Species Abundance Distribution, and various metabolic distributions). Primarily, we consider the discrete distributions predicted by METE. These derivations are not provided in their entirety in Harte (2011)[10], but are the derivations that will produce the core distributions of the Spatial Structure Function and the Ecosystem Structure Function as presented in that work.

In the process of constructing the core structure functions, we derive the Lagrange multipliers that arise from the MaxEnt process, and characterize the ecosystems modeled by METE. We investigate the parameter space of these Lagrange multipliers, and evaluate some of the simplifying assumptions that have been used previously to estimate the Lagrange multiplier values.

We aim to provide a comprehensive resource for ecologists who want to understand the derivations and assumptions of basic structure of METE. We hope that by providing explicit derivations of METE, we will encourage other ecologists to modify the framework, apply it to their own systems, and make progress in this valuable area of research.

## 34 2 | INFORMATION ENTROPY MAXIMIZATION: A PRIMER

35 In this section we present equations that are necessary for information entropy maximization. We then use  
 36 these equations to derive the form of the probability distribution that will result for the simplest case of a discrete,  
 37 1-dimensional distribution. This chain of logic will be applied to the constraints that characterize METE in subsequent  
 38 sections.

### 39 2.1 | Writing down the constraints

40 As observed by Haegeman and Etienne (2010) [21], probability distributions with higher information entropy encode  
 41 less information. Therefore, a probability distribution that corresponds to empirical data without imputing any additional  
 42 information will maximize information entropy. This is also true of a probability distribution that conforms to a  
 43 constraint (as, for example, a constraint on the value of the mean) without making additional assumptions or adding  
 44 other information. In this sense, maximum information entropy methods give the most impartial estimate of the shape of  
 45 the underlying probability distribution for an observable. MaxEnt also gives the least biased estimators of the moments  
 46 of a distribution (which include the range, mean, and variance) [22], meaning that there is no difference, for example,  
 47 between the estimated mean and the empirical mean of a distribution. This feature of MaxEnt is by design, where the  
 48 moments of a probability distribution are constrained by the empirical values of those moments.

49 Here we present the primary equations that regularly occur in information entropy maximization. The general  
 50 expression for  $k$  constraints on the mean value of a distribution of interest is expressed as:

$$\sum_{n=1}^{n=N} f_k(n)p(n) = \langle f_k(n) \rangle \quad (1)$$

51 A second constraint provides for the normalization of the probability distributions, and is expressed as:

$$\sum_{n=1}^{n=N} p(n) = 1 \quad (2)$$

52 The procedure of maximizing entropy (discussed in Box 5.1 of Harte, 2011) results in a particular form of the  
 53 probability distribution and partition function, proven by Jaynes [22] to result in the least-biased probability  $p(n)$  that  
 54 satisfies the "constraint equation" of Eq. (1) and the normalization constraint of Eq. (2),

$$p(n) = \frac{1}{Z} e^{-\sum_{k=1}^{k=K} \lambda_k f_k(n)} \quad (3)$$

55 where  $Z$  is the partition function that serves to normalize the probability distribution, and is expressed as,

$$Z = \sum_{n=1}^{n=N} e^{-\sum_{k=1}^{k=K} \lambda_k f_k(n)} \quad (4)$$

56 Generally, when one wants to use the tools of MaxEnt, one will need to have data from which constraints on the  
 57 distributions can be inferred (such as average values). Then a process of algebra and numerics will begin during which  
 58 one solves for the Lagrange multipliers  $\lambda_k$ . Once the Lagrange multipliers have been determined, then the form of  
 59 the probability distributions can be inferred (or graphed). For examples using Eqs. (1)-(4) see Appendix A.1. Eqs. (3)  
 60 and (4) are written here for reference, as we will use them repeatedly. In the next section we derive Eqs. (3) and (4) by  
 61 maximizing the Shannon information entropy.

## 62 2.2 | The method of Lagrange multipliers and optimization

63 Here we derive the generic probability distribution  $p(n)$  for several specific scenarios of a discrete variable with one  
 64 constraint (beyond normalization), and a discrete variable with multiple constraints. These will serve as simpler examples  
 65 for deriving the core distributions, or structure functions of METE.

66 What does it mean for us to “maximize fairness” using variational calculus? While the explicit form of the probability  
 67 distribution is unknown, we have information about some of its properties. We know that no matter what the form  
 68 of the probability distribution is, it should be properly normalized (that is, the sum of the probabilities of all possible  
 69 outcomes equals to one). This gives us our zeroth constraint, which for a single discrete variable takes the form of

$$\sum_{n=N_{min}}^{N_{max}} p(n) = 1 \quad (5)$$

70 Next, we will likely have some information about an aggregated measurement of the variable in our system. This  
 71 information constitutes our additional constraint or constraints. Letting  $f(n)$  represent a measured value corresponding  
 72 to  $n$ , and assuming that the aggregated measurement we have is the mean value of  $f(n)$ , then we can express our  
 73 knowledge regarding this measurement of an observable quantity that represents some aggregated or average value  
 74 using the mathematical definition of an average.

$$\sum_{n=N_{min}}^{N_{max}} f(n)p(n) = \langle f(n) \rangle \quad (6)$$

75 Thus, the definition itself constitutes our constraint.

76 Lastly, the final piece of information that we have is the functional form of “fairness”, the quantity which we want to  
 77 maximize subject to the above mentioned constraints. This quantity is expressed as the Shannon form of information  
 78 entropy [23]:

$$I = - \sum_{n=N_{min}}^{N_{max}} p(n) \ln(p(n)) \quad (7)$$

79 To maximize “fairness” subject to our constraints, we employ the tools of variational calculus and the method of

undetermined Lagrange multipliers. A concrete example of this is given in Appendix A.1. We begin by constructing the function  $\mathbb{F}$ , which is an expression that incorporates the maximization of information entropy and the constraints we want to incorporate. The constraints are written in such a way so that each constraint is independent of the other, and each term inside the square brackets is zero.

$$\mathbb{F} = - \sum_{n=N_{min}}^{N_{max}} p(n) \ln(p(n)) + \lambda_0 \left[ \sum_{n=N_{min}}^{N_{max}} p(n) - 1 \right] + \lambda_1 \left[ \sum_{n=N_{min}}^{N_{max}} f(n)p(n) - \langle f(n) \rangle \right] \quad (8)$$

In this way, we can incorporate multiple constraints without changing the overall value of the equation. When we perform the “extremization” step to find local optima, the presence of the constraints will change the overall distribution of the probabilities. Maximizing  $\mathbb{F}$  will subsequently maximize “fairness” subject to the additional constraints, thus we seek to solve for the  $p(n)$  that results from  $\partial\mathbb{F}/\partial p(n) = 0$ . Differentiating  $\mathbb{F}$  and setting it equal to zero results in,

$$0 = -[\ln(p(n)) + 1] + \lambda_0(1) + \lambda_1 f(n) \quad (9)$$

Note that we can drop the summations at this point. Solving for  $p(n)$  yields,

$$p(n) = k \exp \{ \lambda_1 f(n) \} \quad (10)$$

where  $k = \exp \{ \lambda_0 - 1 \}$ . Imposing our normalization constraint  $\sum_n p(n) = 1$ , we have

$$\sum_n k \exp \{ \lambda_1 f(n) \} = 1 \quad (11)$$

Since  $k$  is independent of  $n$ , we can factor it out of the summation and solve for it to find,

$$k = \frac{1}{\sum_n \exp \{ \lambda_1 f(n) \}} \quad (12)$$

The expression  $\sum_n \exp \{ \lambda_1 f(n) \}$  comes up so often that it is given its own variable representation  $Z$ , the partition function, which will eventually just turn out to be a real-valued number,

$$Z = \sum_n \exp \{ \lambda_1 f(n) \} \quad (13)$$

Thus, we can express the probability  $p(n)$  as,

$$p(n) = \frac{1}{Z} \exp \{ \lambda_1 f(n) \} = \frac{\exp \{ \lambda_1 f(n) \}}{\sum_n \exp \{ \lambda_1 f(n) \}} \quad (14)$$

To find the explicit form of  $p(n)$ , one needs further information in the form of data. Having data, it is then possible to use the constraint equations to numerically solve for the undetermined Lagrange multipliers, and thus identify the form of the probability distribution  $p(n)$ .

### 97 3 | THE STRUCTURE OF METE

98 In this section, we introduce and discuss the ecological state variables common to macroecology, as used by METE.  
99 We apply the MaxEnt method to these ecological state variables to derive the core distributions of METE, namely the  
100 Spatial Structure Function, and the Ecosystem Structure Function [10, 11].

#### 101 3.1 | A state variable theory

102 Much of macroecology is concerned with detecting patterns in ecosystems, either at large scale, or as an emergent  
103 property of scaling over larger aggregates of individuals, species, area, or time [24]. To detect these patterns, we  
104 must work with variables that are sufficiently coarse that they capture average conditions of some larger phenomena  
105 that is being modeled, or sufficiently aggregated that they apply across systems and capture certain generalities. For  
106 example, we may examine average abundance of all individuals in an area. The “average” may refer to an average of  
107 repeated measurements in different plots or subplots, for example, which can smooth out heterogeneity and model  
108 average conditions on a landscape. The “aggregated” nature of abundance refers to something slightly different, in  
109 that abundance measured will be a result of multiple processes, such as birth, death, emigration, and immigration [25].  
110 Abundance therefore represents an “aggregate” metric of all of these processes. Similarly, “species” as an observable may  
111 have an average value when measured in many similar sized plots, but is itself an aggregate measure of biodiversity that  
112 ignores (or “coarse grains”) genetic variation among individuals, and processes such as hybridization. These variables  
113 are easily measured and compared across ecological systems, and have an overall generality and transferability that  
114 make them interesting at large scales.

115 Within METE, variables representing total area, total number of species, total abundance, and total metabolic  
116 energy of an ecological system are central to the definitions of the core probability distributions. These ecological  
117 state variables are represented as  $A_0$ ,  $S_0$ ,  $N_0$ , and  $E_0$ , respectively. They are static (not time-dependent), and can be  
118 used to model macroecological distributions of interest, such as the Species-Area Relationship, the Species Abundance  
119 Distribution, and various metabolic rate distributions, both within a species and across an entire community. The “ASNE”  
120 version of METE has been the subject of the most study [10], but other constraints featuring additional state variables  
121 are possible. In one case, METE has been extended to include higher taxonomic constraints [26].

122 In the following sections, we will demonstrate how the state variables and their ratios are used in the MaxEnt  
123 procedure to constrain patterns of individuals over area for the Spatial Structure Function, and patterns of metabolic  
124 requirement (or body size) across individuals and abundances per species through the Ecosystem Structure Function.

#### 125 3.2 | The Spatial Structure Function

126 This distribution goes by multiple names among practitioners, including the “Pi distribution” (informally), the “Species-  
127 level spatial abundance distribution” or SSAD [14, 27], and the Spatial Structure Function (SSF) [28], which is how we  
128 will refer to it here. The SSF is a by-species prediction of the clustering of individuals over space, and is defined as the  
129 “probability that  $n$  individuals of a species are found in a cell of area  $A$  if it has  $n_0$  individuals in  $A_0$ ” [10]. We derive this  
130 distribution before the Ecosystem Structure Function for pedagogical reasons, as it is a one-dimensional distribution for  
131 a discrete variable. This variable is the the abundance of a single species,  $n$ , at a given scale,  $A$ , where  $A$  is a smaller area  
132 within the total area under consideration,  $A_0$ . The variables that are used to constrain the system are  $A$ ,  $A_0$ , and the  
133 abundance of a single species at the total spatial scale,  $n_0$ . The derivation of the spatial structure function will follow  
134 exactly the general approach provided in the previous section. We represent this function as  $\Pi(n|A, n_0, A_0)$ . To ensure

135 that the spatial structure function is properly normalized, we express our zeroth constraint as,

$$\sum_{n=0}^{n_0} \Pi(n) = 1 \quad (15)$$

136 where we have dropped the conditional variables  $A$ ,  $n_0$ , and  $A_0$  for notational convenience. Note also that the lower  
137 limit on the summation is set to  $n = 0$ . This is because  $n$  represents the per-species abundance, and it is possible for a  
138 species that has  $n_0$  individuals to have zero abundance in an area  $A$ .

139 Our additional constraint comes from the simultaneous definition and measurement of the average value of the  
140 per-species abundance  $\bar{n}$ . From measurement,  $\bar{n} = n_0 A / A_0$ . From definition,  $\bar{n} = \sum_0^{n_0} n \Pi(n)$ . Combining these two  
141 expressions gives us our other constraint,

$$\sum_{n=0}^{n_0} n \Pi(n) = \frac{n_0 A}{A_0} \quad (16)$$

142 Now, as we want to maximize “fairness”, or information entropy related to the spatial structure function subject to the  
143 above two constraints, we construct the function  $\mathbb{F}$  in the form,

$$\mathbb{F} = - \sum_{n=0}^{n_0} \Pi(n) \ln(\Pi(n)) - \lambda_0 \left[ \sum_{n=0}^{n_0} \Pi(n) - 1 \right] - \lambda_{\Pi} \left[ \sum_{n=0}^{n_0} n \Pi(n) - \frac{n_0 A}{A_0} \right] \quad (17)$$

144 From here we maximize  $\mathbb{F}$  by evaluating the expression  $\partial \mathbb{F} / \partial \Pi(n) = 0$ . This results in,

$$0 = -[\ln(\Pi(n)) + 1] - \lambda_0(1) - \lambda_{\Pi}n \quad (18)$$

145 From here we can solve for  $\Pi(n)$  to arrive at,

$$\Pi(n) = k \exp\{-\lambda_{\Pi}n\} \quad (19)$$

146 where  $k = \exp\{-(\lambda_0 + 1)\}$ . Imposing our zeroth normalization constraint, we can write,

$$\sum_{n=0}^{n_0} k \exp\{-\lambda_{\Pi}n\} = 1 \quad (20)$$

147 Since  $k$  is independent of  $n$ , we can factor it out of the summation and rewrite as,

$$k = \frac{1}{\sum_{n=0}^{n_0} \exp\{\lambda_{\Pi}n\}} \quad (21)$$

148 Conventionally this normalization constant is defined as  $1/Z$  and denoted as the partition function, where,

$$Z = \sum_{n=0}^{n_0} \exp\{-\lambda_{\Pi}n\} \quad (22)$$

149 This brings us to the following expression compact expression for the Spatial Structure Function,

$$\Pi(n) = \frac{1}{Z} \exp\{-\lambda_{\Pi} n\} \quad (23)$$

150 In order to actually calculate, or graph, the spatial structure function for a given set of values  $A_0$ ,  $n_0$ , and  $A$ , we must  
 151 first calculate the Lagrange multiplier  $\lambda_{\Pi}$  using our constraint equation relating the measured average per-species  
 152 abundance  $\bar{n} = n_0 A / A_0$  to the definition of the average per-species abundance  $\bar{n} = \sum_{n=0}^{n_0} n \Pi(n)$ . This time, substituting  
 153 our known expression for  $\Pi(n)$ , we have,

$$\sum_{n=0}^{n_0} \frac{n}{Z} \exp\{-\lambda_{\Pi} n\} = \frac{n_0 A}{A_0} \quad (24)$$

154 Recalling that  $Z$  is independent of  $n$ , it can be pulled out of the summation. Furthermore, substituting our definition of  
 155  $Z$ , but using  $m$  as a dummy index instead of  $n$  to avoid mixing up our indices, we can write,

$$\frac{\sum_{n=0}^{n_0} n \exp\{-\lambda_{\Pi} n\}}{\sum_{m=0}^{n_0} \exp\{-\lambda_{\Pi} m\}} = \frac{n_0 A}{A_0} \quad (25)$$

156 In principle, the above expression allows one to solve for the Lagrange multiplier  $\lambda_{\Pi}$ . However, analytical solutions for  
 157  $\lambda_{\Pi}$  are intractable, thus one must resort to numerical methods. We graph the parameter space of the state variables  
 158  $A_0$ ,  $n_0$ ,  $n$  and  $\lambda_{\Pi}$  in Figure 1.

### 159 3.3 | The Ecosystem Structure Function

160 The Ecosystem Structure Function (ESF) is the second of METE's core distributions. Unlike the SSF, it does not have  
 161 a simple definition, but can be described as a kind of "container function" that describes the probability space of how  
 162 abundances are assigned to species and how metabolic energy is partitioned over individuals in a community. The  
 163 ESF is the distribution from which the Species Abundance Distribution, Species Area Relationship, and Metabolic Rate  
 164 Distributions [14, 18] can eventually be derived. Constrained by empirical values measured from real systems, the ESF  
 165 produces actual predictions of these probability distributions. For an in-depth discussion of the definition of the METE  
 166 ESF, see Bertram et al. (2019) [29].

167 As the ESF  $R(n, \epsilon)$ , depends on one discrete variable  $n$ , and one continuous variable  $\epsilon$ , we will this time need to  
 168 integrate of  $\epsilon$  in addition to summing over  $n$ . Thus, our normalization constraint now takes the form of,

$$\sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} R(n, \epsilon) d\epsilon = 1 \quad (26)$$

169 Our additional constraints are aggregated measures of variables  $n$  and  $n\epsilon$ , that is  $f_1(n) = n$ , and  $f_2(n\epsilon) = n\epsilon$ . The  
 170 measures themselves are the ratios  $N_0/S_0$  and  $E_0/S_0$ , that is  $\langle f_1(n) \rangle = N_0/S_0$ , and  $\langle f_2(n\epsilon) \rangle = E_0/S_0$ . This gives us the pair

171 of constraints,

$$\sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n R(n, \epsilon) d\epsilon = \frac{N_0}{S_0} \quad (27)$$

$$\sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \epsilon R(n, \epsilon) d\epsilon = \frac{E_0}{S_0} \quad (28)$$

172 Now, the function  $\mathbb{F}$  that we will want to maximize takes the form of,

$$\begin{aligned} \mathbb{F} = & - \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} R(n, \epsilon) \ln(R(n, \epsilon)) d\epsilon - \lambda_0 \left[ \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} R(n, \epsilon) d\epsilon - 1 \right] \\ & - \lambda_1 \left[ \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n R(n, \epsilon) d\epsilon - \frac{N_0}{S_0} \right] - \lambda_2 \left[ \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \epsilon R(n, \epsilon) d\epsilon - \frac{E_0}{S_0} \right] \end{aligned} \quad (29)$$

173 As before, we maximize  $\mathbb{F}$  by evaluating  $\partial \mathbb{F} / \partial R = 0$ . This results in,

$$0 = -[\ln(R(n, \epsilon)) + 1] - \lambda_0 [1] - \lambda_1 [n] - \lambda_2 [n\epsilon] \quad (30)$$

174 Solving for  $R(n, \epsilon)$  gives,

$$R(n, \epsilon) = k \exp\{-\lambda_1 n - \lambda_2 n\epsilon\} \quad (31)$$

175 where  $k = \exp\{-(1 + \lambda_0)\}$ . Using our normalization condition to define  $Z$  we have,

$$Z = \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} \exp\{-\lambda_1 n - \lambda_2 n\epsilon\} d\epsilon \quad (32)$$

176 This allows us to express the full form of the ecosystem structure function as,

$$R(n, \epsilon) = \frac{\exp\{-\lambda_1 n - \lambda_2 n\epsilon\}}{\sum_{m=1}^{N_0} \int_{\epsilon'=1}^{E_0} \exp\{-\lambda_1 m - \lambda_2 m\epsilon'\} d\epsilon'} \quad (33)$$

177 where we have replaced  $n$  with  $m$  and  $\epsilon$  and  $\epsilon'$  in the denominator to ensure there is no confusion over which variables  
178 belong in the numerator or denominator.

179 At this point we can simplify the expression for  $Z$  by performing the integral over  $\epsilon'$ . Factoring out the term in  $Z$   
180 independent of  $\epsilon'$  we have,

$$Z = \sum_{m=1}^{N_0} \exp\{-\lambda_1 m\} \int_{\epsilon'=1}^{E_0} \exp\{-\lambda_2 m\epsilon'\} d\epsilon' \quad (34)$$

181 After integrating, we have

$$Z = \sum_{m=1}^{N_0} \frac{\exp\{-\lambda_1 m\}}{\lambda_2 m} [\exp\{-\lambda_2 m\} - \exp\{\lambda_2 m E_0\}] \quad (35)$$

182 And with further simplification,

$$Z = \frac{1}{\lambda_2} \sum_{m=1}^{N_0} \frac{1}{m} [\exp\{-m\beta\} - \exp\{-m\sigma\}] \quad (36)$$

183 where  $\beta = \lambda_1 + \lambda_2$  and  $\sigma = \lambda_1 + E_0 \lambda_2$ .

184 From here we can examine the constraint Eqs. 27 and 28. In particular, now that we have an explicit form for the  
 185 structure function, we can perform the integrals in the constraint equations to fix the values of the unknown Lagrange  
 186 multipliers in terms of the measured quantities  $N_0$ ,  $E_0$ , and  $S_0$ . Upon substitution of our expression for  $R(n, \epsilon)$  into Eq.  
 187 27 we have,

$$\frac{N_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \exp\{-n(\lambda_1 + \lambda_2 \epsilon)\} d\epsilon \quad (37)$$

188 Factoring from the integral the term independent of  $\epsilon$ ,

$$\frac{N_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} n \exp\{-n\lambda_1\} \int_{\epsilon=1}^{E_0} \exp\{-n\lambda_2 \epsilon\} d\epsilon \quad (38)$$

189 Upon integration,

$$\frac{N_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} n \exp\{-n\lambda_1\} \left[ \frac{\exp\{-n\lambda_2\} - \exp\{-n\lambda_2 E_0\}}{n\lambda_2} \right] \quad (39)$$

190 And with further simplification,

$$\frac{N_0}{S_0} = \frac{1}{Z \lambda_2} \sum_{n=1}^{N_0} [\exp\{-n\beta\} - \exp\{-n\sigma\}] \quad (40)$$

191 where  $\beta = \lambda_1 + \lambda_2$  and  $\sigma = \lambda_1 + E_0 \lambda_2$ .

192 Turning our attention now to constraint Eq. 28, upon substitution of  $R(n, \epsilon)$  we have,

$$\frac{E_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} \int_{\epsilon=1}^{E_0} n \epsilon \exp\{-n(\lambda_1 + \lambda_2 \epsilon)\} d\epsilon \quad (41)$$

193 Factoring from the integral the term independent of  $\epsilon$ ,

$$\frac{E_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} n \exp\{-n\lambda_1\} \int_{\epsilon=1}^{E_0} \epsilon \exp\{-n\lambda_2 \epsilon\} d\epsilon \quad (42)$$

194 After integrating by parts we have,

$$\frac{E_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} n \exp\{-n\lambda_1\} \left[ \frac{\exp\{-n\lambda_2\} - E_0 \exp\{-n\lambda_2 E_0\}}{n\lambda_2} + \frac{\exp\{-n\lambda_2\} - \exp\{-n\lambda_2 E_0\}}{(n\lambda_2)^2} \right] \quad (43)$$

195 And with further simplification,

$$\frac{E_0}{S_0} = \frac{1}{Z} \sum_{n=1}^{N_0} \left[ \frac{\exp\{-n\beta\} - E_0 \exp\{-n\sigma\}}{\lambda_2} + \frac{\exp\{-n\beta\} - \exp\{-n\sigma\}}{n\lambda_2^2} \right] \quad (44)$$

196 where  $\beta = \lambda_1 + \lambda_2$  and  $\sigma = \lambda_1 + E_0\lambda_2$ .

197 Substituting in our expression for  $Z$ , we can write the integrated versions of the constraint equations in full as,

$$\frac{N_0}{S_0} = \frac{\sum_{n=1}^{N_0} [e^{-n\beta} - e^{-n\sigma}]}{\sum_{m=1}^{N_0} \left[ \frac{e^{-m\beta} - e^{-m\sigma}}{m} \right]} \quad (45)$$

$$\frac{E_0}{S_0} = \frac{\sum_{n=1}^{N_0} [e^{-n\beta} - E_0 e^{-n\sigma}]}{\sum_{m=1}^{N_0} \left[ \frac{e^{-m\beta} - e^{-m\sigma}}{m} \right]} + \frac{1}{\lambda_2} \quad (46)$$

198 The values of  $\lambda_1$  and  $\lambda_2$  are often difficult to calculate by conventional means, so some approximations were used in  
 199 Harte (2011). With the use of meteR [28], these approximations are no longer necessary. However, we consider them in  
 200 more depth in Appendix B.

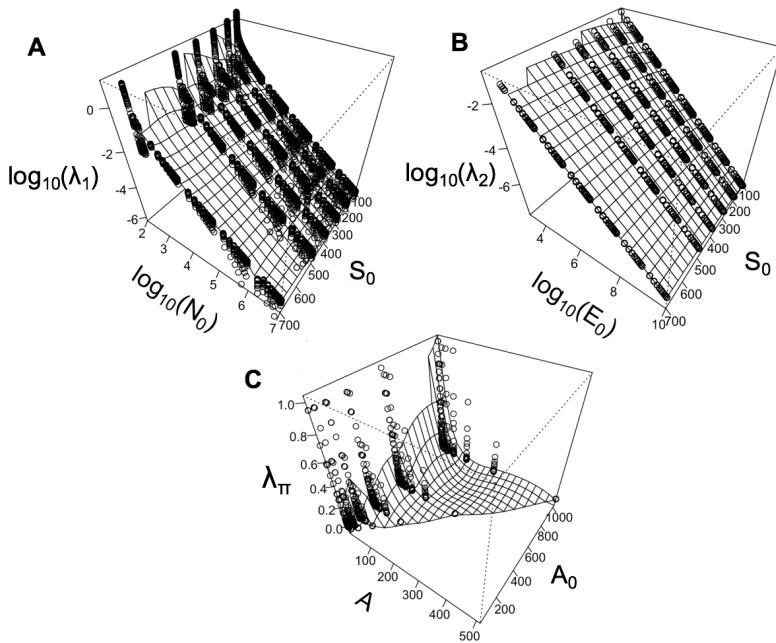
## 201 4 | RELATIONSHIPS BETWEEN STATE VARIABLES AND LAGRANGE MULTI- 202 PLIERS

203 By using the MaxEnt approach with the METE ecological state variables, we derive three Lagrange multipliers:  $\lambda_{\Pi}$   
 204 associated with the Spatial Structure Function, and  $\lambda_1$  and  $\lambda_2$ , associated with the Ecosystem Structure Function. We  
 205 graph these, along with the ecological state variables in their respective constraints, in Figure 1. Examining  $\lambda_1$  in panel  
 206 (A), we see the greater influence of  $\log(N_0)$  than  $S_0$  on the overall value of the Lagrange multiplier  $\lambda_1$ , and a compression  
 207 of  $\lambda_1$  values at low  $N_0$ . This may correspond to METE's own stated limitations, and its requirement that  $N \gg 1$  [10]. In  
 208 panel (B), we can see a near-linear relationship on the log-log scale between  $\lambda_2$  and  $\log(E_0)$ , while  $S_0$  does not affect  
 209 its value as greatly over this range of values. In panel (C),  $\lambda_{\Pi}$  is clearly non-linear in both the state variable  $A_0$ , and the  
 210 smaller area under consideration,  $A$ .

211 The METE Lagrange multipliers represent all the possible relationships between  $N_0$ ,  $S_0$ , and  $E_0$ , and METE's  
 212 predicted relationships of  $n$ ,  $A$ , and  $A_0$ . The graph of  $\lambda_{\Pi}$  is interesting because it appears that very different values  
 213 of  $\lambda_{\Pi}$  may be obtained by varying the ratio of  $A$  to  $A_0$  slightly, and this may in turn suggest that it does not have the  
 214 properties we would desire in a metric describing clustering. It has been demonstrated that the SSF does not always  
 215 produce reliable predictions for clustering of individuals of a species within a given area [16, 27], and this area of METE  
 216 could be extended and modified in future work. Further investigations of the relationships of  $\lambda_1$  and  $\lambda_2$  with data from  
 217 varied ecosystems may also allow us to investigate patterns of diversity, abundance, body size, and the relationships  
 218 between these macroecological variables in new ways. These kinds of parameter space representations are also useful  
 219 in generating hypotheses about changing ecosystems, and what new distributions are expected as one or more state

220 variable changes.

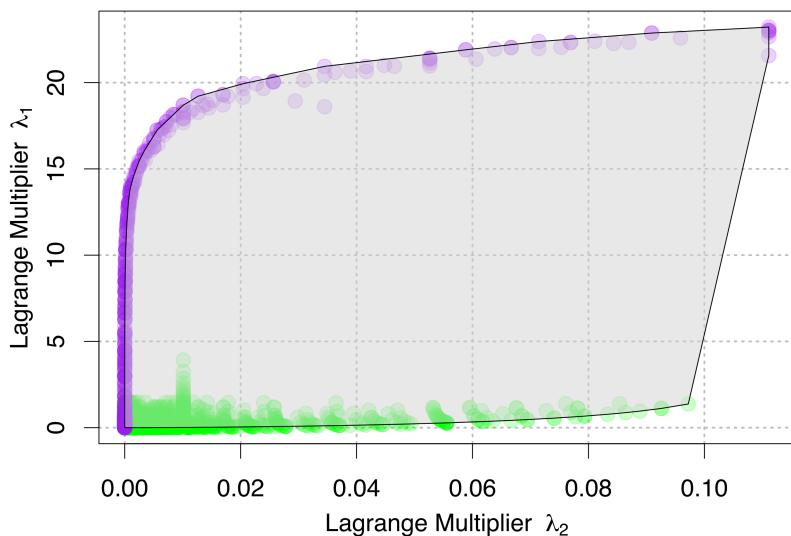
221 In Figure 2, we graph the parameter space that is defined by the ESF through the Lagrange multipliers  $\lambda_1$  and  $\lambda_2$ . In  
 222 this graph, the boundaries of the defined parameter space become interesting. High values of  $\lambda_1$  always correspond to  
 223 one or more “singleton species,” or species with a single individual, where  $N_0 = S_0$ . This can only happen when a single  
 224 individual is measured in order to estimate the values of  $A_0$ ,  $N_0$ ,  $S_0$ , and  $E_0$  (that is, small numbers of measurements), or  
 225 in cases where there are the most species possible given the number of individuals present (extreme diversity). These  
 226 cases therefore represent theoretical limits of possible outcomes of measurement. We expect that most real systems  
 227 will have many measurements of species with more than one individual, and will fall into the range of low  $\lambda_1$  values. The  
 228 behavior of the lower values of the graph may therefore be worth investigating further.



**FIGURE 1** The relationship between the METE’s Lagrange multipliers  $\lambda_1$ ,  $\lambda_2$ , and  $\lambda_\pi$ , and the ecological state variables in the mathematical constraints that produce them. Values for each  $\lambda$  were generated with meteR [28], and a surface was interpolated to aid in visualization. In panel (A), we see the greater influence of  $\log(N_0)$  than  $S_0$  on the overall value of the Lagrange multiplier  $\lambda_1$ , and a compression of  $\lambda_1$  values at low  $N_0$ . In panel (B), we can see a near-linear relationship on the log-log scale between  $\lambda_2$  and  $\log(E_0)$ , while  $S_0$  does not affect its value as greatly over this range of values. In panel (C), we see a highly non-linear relationship between  $\lambda_\pi$ , the state variable  $A_0$ , and the smaller area under consideration,  $A$ .

## 229 5 | SUMMARY

230 In this paper, we provide an in-depth look at the mathematical derivations underlying the Maximum Entropy Theory of  
 231 Ecology, or METE. These derivations fill a gap for those researchers who would like to see the explicit chain of logic that  
 232 produces the core structure functions of METE, from which the predictions of METE follow. In presenting this theory,



**FIGURE 2** The parameter space of ecosystems as defined by the METE Lagrange multipliers  $\lambda_1$ , corresponding to the constraint on  $\frac{N_0}{S_0}$ , and  $\lambda_2$ , corresponding to the constraint on  $\frac{E_0}{S_0}$ . The highest values of  $\lambda_1$  for any value of  $\lambda_2$  correspond to values of  $\frac{N_0}{S_0} = 1$  (shown in purple), or situations where there is only one individual per species (small numbers of measurements or high diversity). Most real ecosystems and empirical systems with more than a few individuals are expected to fall closer to the low  $\lambda_1$  values for any given  $\lambda_2$  value (shown in green).

233 we take a different pedagogical approach than is employed in Harte (2011) [10]. Namely, we provide a general case of a  
 234 discrete MaxEnt problem in one dimension. We then give worked examples of constraints on moments beyond just the  
 235 mean of a distribution (in Appendix A). From there, we first construct the simpler Spatial Structure Function, which is a  
 236 discrete probability distribution, and then work through the logic of the Ecosystem Structure Function, which contains  
 237 more constraints, and has a combination of discrete and continuous variables.

238 We do not cover the applications of METE to testing theory against empirical data, which has been done extensively  
 239 elsewhere [9, 12, 10, 13, 11, 14, 15, 16, 17, 18, 19, 20], however, we explicitly show the MaxEnt process and how it is  
 240 applied to the ecological state variables of METE to produce the core structure functions from which all the predictions  
 241 of METE derive.

242 Finally, we discuss the three Lagrange multipliers that result from applying the MaxEnt procedure to METE's  
 243 ecological state variables, and how these Lagrange multipliers characterize the system being studied. We investigate (in  
 244 Appendix B) some of the simplifying approximations that were previously used to evaluate the Lagrange multipliers and  
 245 assess their realism.

246 The equations presented here should make it easier for other researchers to make advances in MaxEnt-based  
 247 macroecology, and METE in particular, either by investigating new state variables, or new functional forms of the  
 248 mathematical constraints. The Lagrange multipliers, graphed for the first time here, form a parameter space that may be  
 249 useful to further studies involving comparisons between ecosystems and across scales.

---

**250 ACKNOWLEDGEMENTS**

251 We thank J. Harte, V. Savage, D. Hembry, D. McKenzie, J. Bertram, and B. Enquist for discussions and edits relevant to  
252 this manuscript.

**253 CONFLICT OF INTEREST**

254 The authors declare no conflicts of interest.

**255 REFERENCES**

256 [1] Brown JH. *Macroecology*. University of Chicago Press; 1995.

257 [2] McGill BJ, Etienne RS, Gray JS, Alonso D, Anderson MJ, Bencha H, et al. Species abundance distributions: moving be-  
258 yond single prediction theories to integration within an ecological framework. *Ecology letters* 2007;10(10):995–1015.

259 [3] Yoda K. Self-thinning in overcrowded pure stands under cultivated and natural conditions (Intraspecific competition  
260 among higher plants. XI). *J Inst Polytch Osaka City Univ Ser D* 1963;14:107–129.

261 [4] Damuth J. Population density and body size in mammals. *Nature* 1981;290(5808):699.

262 [5] Enquist BJ, Niklas KJ. Invariant scaling relations across tree-dominated communities. *Nature* 2001;410(6829):655.

263 [6] West GB, Brown JH, Enquist BJ. A general model for the origin of allometric scaling laws in biology. *Science*  
264 1997;276(5309):122–126.

265 [7] Gillooly JF, Brown JH, West GB, Savage VM, Charnov EL. Effects of size and temperature on metabolic rate. *science*  
266 2001;293(5538):2248–2251.

267 [8] Brown JH, Gillooly JF, Allen AP, Savage VM, West GB. Toward a metabolic theory of ecology. *Ecology* 2004;85(7):1771–  
268 1789.

269 [9] Harte J, Zillio T, Conlisk E, Smith AB. Maximum entropy and the state-variable approach to macroecology. *Ecology*  
270 2008;89(10):2700–2711.

271 [10] Harte J. *Maximum entropy and ecology: a theory of abundance, distribution, and energetics*. Oxford University Press;  
272 2011.

273 [11] Harte J, Newman EA. Maximum information entropy: a foundation for ecological theory. *Trends in ecology & evolution*  
274 2014;29(7):384–389.

275 [12] Harte J, Smith AB, Storch D. Biodiversity scales from plots to biomes with a universal species-area curve. *Ecology letters*  
276 2009;12(8):789–797.

277 [13] White EP, Thibault KM, Xiao X. Characterizing species abundance distributions across taxa and ecosystems using a sim-  
278 ple maximum entropy model. *Ecology* 2012;93(8):1772–1778.

279 [14] Newman EA, Harte ME, Lowell N, Wilber M, Harte J. Empirical tests of within-and across-species energetics in a diverse  
280 plant community. *Ecology* 2014;95(10):2815–2825.

281 [15] Harte J, Kitzes J. Inferring regional-scale species diversity from small-plot censuses. *PLoS one* 2015;10(2):e0117527.

282 [16] McGlinn D, Xiao X, Kitzes J, White EP. Exploring the spatially explicit predictions of the Maximum Entropy Theory of  
283 Ecology. *Global Ecology and Biogeography* 2015;24(6):675–684.

284 [17] Wilber MQ, Kitzes J, Harte J. Scale collapse and the emergence of the power law species-area relationship. *Global*  
285 *ecology and biogeography* 2015;24(8):883–895.

286 [18] Xiao X, McGinn DJ, White EP. A strong test of the maximum entropy theory of ecology. *The American Naturalist*  
287 2015;185(3):E70–E80.

288 [19] Harte J, Newman EA, Rominger AJ. Metabolic partitioning across individuals in ecological communities. *Global Ecology*  
289 *and Biogeography* 2017;26(9):993–997.

290 [20] Shoemaker WR, Locey KJ, Lennon JT. A macroecological theory of microbial biodiversity. *Nature ecology & evolution*  
291 2017;1(5):0107.

292 [21] Haegeman B, Etienne RS. Entropy maximization and the spatial distribution of species. *The American Naturalist*  
293 2010;175(4):E74–E90.

294 [22] Jaynes ET. Information theory and statistical mechanics. *Physical review* 1957;106(4):620.

295 [23] Shannon CE. A mathematical theory of communication. *Bell system technical journal* 1948;27(3):379–423.

296 [24] McGill BJ. The what, how and why of doing macroecology. *Global Ecology and Biogeography* 2019;28(1):6–17.

297 [25] Hubbell SP. The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press; 2001.

298 [26] Harte J, Rominger A, Zhang W. Integrating macroecological metrics and community taxonomic structure. *Ecology letters*  
299 2015;18(10):1068–1077.

300 [27] Newman EA, Wilber MQ, Kopper KE, Moritz MA, Falk DA, McKenzie D, et al. Disturbance macroecology: integrating  
301 disturbance ecology and macroecology in different-age post-fire stands of a closed-cone pine forest as a case study. *BioRxiv*  
302 2018;p. 309419.

303 [28] Rominger AJ, Merow C. meteR: an r package for testing the maximum entropy theory of ecology. *Methods in Ecology*  
304 *and Evolution* 2017;8(2):241–247.

305 [29] Bertram J, Newman EA, Dewar R. Maximum entropy models elucidate the contribution of metabolic traits to patterns  
306 of community assembly. *bioRxiv* 2019;p. 526764.

307 [30] Enquist BJ, Norberg J, Bonser SP, Violette C, Webb CT, Henderson A, et al. Scaling from traits to ecosystems: developing a  
308 general trait driver theory via integrating trait-based and metabolic scaling theories. In: *Advances in ecological research*,  
309 vol. 52 Elsevier; 2015.p. 249–318.

310 [31] Wieczynski DJ, Boyle B, Buzzard V, Duran SM, Henderson AN, Hulshof CM, et al. Climate shapes and shifts functional  
311 biodiversity in forests worldwide. *Proceedings of the National Academy of Sciences* 2019;116(2):587–592.

312 **SUPPLEMENTARY MATERIAL**313 **A | EXAMPLES APPLYING MAXENT TO KNOWN DISTRIBUTIONS**314 **A.1 | A Fair 3-Sided Die Constrained by the Mean**

315 Consider the case of a 3-sided die with sides labelled 1, 2, and 3. Suppose we perform an experiment (that is, rolling a  
 316 die), and measure the result (that is, reading and recording the outcome of the roll in the form of the number on the  
 317 face-up side). After performing enough of these experiments, we can determine that the long-term average value of the  
 318 rolls is to equal 2, by adding up all the outcomes and then dividing by the number of rolls. We take note that this is also  
 319 the *true* average of a fair die, where all outcomes are equally probable. Because we chose a *fair* die in the first place,  
 320 we are not surprised that our empirical outcome matches the outcome for a fair die. However, we are interested in  
 321 determining the probability distribution that is associated with rolling any die, fair or not. This mathematical approach  
 322 will allow us to find our empirical average and figure out what the probabilities of each outcome are, whether or not we  
 323 have a fair die. To work through the logic, we will use the tools presented in the form of equations Eqs. (1, 3, 4).

324 First, we start by writing Eq. [1], and substituting in the values corresponding to our problem. Note that we have  
 325 only one constraint (knowledge of the empirical mean), so  $K = 1$ , and all sums over  $k$  vanish. Thus,  $f_k(n)$  becomes  $f(n)$ .  
 326 Next, we can only make three possible observations, corresponding to the values on the three sides of the dice. So  
 327  $f(1) = 1$ ,  $f(2) = 2$ , and  $f(3) = 3$ . Finally, assuming that we have performed this experiment and have a mean value for  $f(n)$   
 328 to calculate from the possible observations, we can write that  $\langle f(n) \rangle = (3 + 2 + 1)/3 = 2$ . So, Eq. (1) becomes,

$$329 \quad 1p(1) + 2p(2) + 3p(3) = 2 \quad (47)$$

330 This can be read as “the sum of the probabilities of a particular outcome of a roll times the value of that roll equals  
 331 the empirical average of all outcomes on repeated trials.” Next, we can substitute our problem-specific values into the  
 332 definitions of the probability distribution  $p(n)$  and the partition function  $Z$ . Doing so makes Eq. (3) take on the form:

$$333 \quad p(n) = \frac{1}{Z} e^{-\lambda n} \quad (48)$$

332 for generic values of  $n$ , or specifically,

$$333 \quad \begin{aligned} p(1) &= \frac{1}{Z} e^{-\lambda} \\ p(2) &= \frac{1}{Z} e^{-\lambda 2} \\ p(3) &= \frac{1}{Z} e^{-\lambda 3} \end{aligned} \quad (49)$$

334 We have one Lagrange multiplier, here denoted as  $\lambda$ , and the probability of each outcome is expressed in terms  
 335 of the partition function, as well as a term that involves the actual value of the outcome multiplied by the Lagrange  
 336 multiplier. The partition function, which is defined in Eq. (4), takes on the form:

$$Z = \sum_{n=1}^{n=3} e^{-\lambda n}$$

$$Z = e^{-\lambda} + e^{-\lambda 2} + e^{-\lambda 3} \quad (50)$$

336 Remember that in the end, we seek the specific form of the probability distribution function, which means that we  
 337 have to solve for the Lagrange multiplier,  $\lambda$ , that appears in both the probability distribution and the partition function.  
 338 We can now substitute our specified version of the partition function into the specified versions of the probability  
 339 distributions, and substitute those into the specified version of our constraint equation. We will do this step by step  
 340 (and technically in reverse order), starting with substituting the probabilities into the constraint, or Eq. (49) into Eq. (48).  
 341 This gives us,

$$\frac{1}{Z} e^{-\lambda} + \frac{2}{Z} e^{-\lambda 2} + \frac{3}{Z} e^{-\lambda 3} = 2 \quad (51)$$

342 Before substituting our partition function,  $Z$ , Eq. (50), into the constraint equation, Eq. (48), we will first multiply  
 343 both sides by the partition function,  $Z$ , to simplify the expression, giving,

$$e^{-\lambda} + 2e^{-\lambda 2} + 3e^{-\lambda 3} = 2Z \quad (52)$$

344 Now we will substitute in the expression for the partition function,  $Z$ , to arrive at:

$$e^{-\lambda} + 2e^{-\lambda 2} + 3e^{-\lambda 3} = 2(e^{-\lambda} + e^{-\lambda 2} + e^{-\lambda 3}) \quad (53)$$

345 Here we will make an extra substitution, simply for convenience. The equation above is – at heart – a polynomial  
 346 expression in  $e^{-\lambda}$ . So, let's substitute  $e^{-\lambda} = x$ . Now we will be solving for  $x$ , and at the end we will see what this tells us  
 347 about  $\lambda$ . Making this substitution, and performing the necessary algebra, we have,

$$x + 2x^2 + 3x^3 = 2x + 2x^2 + 2x^3$$

$$0 = x - x^3$$

$$0 = x(1 - x^2) \quad (54)$$

348 Thus, we have found that the solutions for  $x$  are  $x = 0$ ,  $x = 1$ , or  $x = -1$ . Recalling that we made the substitution  
 349  $x = e^{-\lambda}$ , this means that we really have  $e^{-\lambda} = 0$ ,  $e^{-\lambda} = 1$ , or  $e^{-\lambda} = -1$ . With these possibilities, we will need to check if  
 350 one of these possible solutions is the true answer. The other two solutions will prove to give non-sensible probability  
 351 distributions (complex numbers or non-physical solutions) for the problem that we have specified. To find out which  
 352 one of the three is correct we must examine the forms of the probability distributions that they yield. We could do  
 353 this one of two ways. We could use natural logarithm rules to determine the exact possible values of  $\lambda$ , and substitute

354 back into the probability distribution and partition function formulas to find their forms. Or, since both the probability  
 355 distribution and partition function formulas are functions of  $e^{-\lambda}$ , we could simply substitute the various known values  
 356 of 0, -1, or 1 for each occurrence of  $e^{-\lambda}$ . We will do the latter, as it involves fewer steps.

357 If  $e^{-\lambda} = 0$  were the correct answer, then the partition function would take the form of  $Z = 0^1 + 0^2 + 0^3 = 0$ , and the  
 358 probability distributions would take the form of  $p(n) = 0^n / Z = 0/0$ . Dividing by zero is clearly a problem, so  $e^{-\lambda} = 0$  is  
 359 not the correct answer. Alternatively, if  $e^{-\lambda} = -1$  were the correct answer, then the partition function would take the  
 360 form of  $Z = -1^1 + (-1)^2 + (-1)^3 = -1$ , and the probability distribution would take the form of  $p(n) = (-1)^n / -1 = -(-1)^n$ .  
 361 This is also a problem, as it means that every probability for rolling an even number would be negative! So  $e^{-\lambda} = -1$  is  
 362 also not the correct answer. This means that  $e^{-\lambda} = 1$  must be the correct answer. In this case, the partition function takes  
 363 the form of  $Z = 1^1 + 1^2 + 1^3 = 3$ , and the probability distribution takes the form of  $p(n) = 1^n / 3 = 1/3$ . Remembering  
 364 that we have a *fair* die, where *fair* means equal chances of rolling any number, we see that we have indeed found the  
 365 correct answer.

## 366 A.2 | A Fair 3-Sided Die Constrained by the Standard Deviation

367 Here we consider a case similar to that above, but with a minor variation. We will constrain the problem with knowledge  
 368 regarding the standard deviation of the rolls of the die, instead of knowledge regarding the mean value of the rolls  
 369 of the die. It is still a fair three-sided die, and we will still use the *true* value one would get for the standard deviation  
 370 of actual rolls so that we can check our answer against reality at the end. This problem is of interest to us for several  
 371 reasons. First, it is possible to contrive a situation where all you may know about your data is the standard deviation,  
 372 and not the mean, yet you still want to determine the corresponding form of the probability distribution functions  
 373 (perhaps you are reconstructing data from faded paper articles that originally reported both the mean and the standard  
 374 deviation). Second, As a recent example of the importance of understanding the moments of distributions, trait-driver  
 375 theory (TDT) in macroecology suggests potential applications using MaxEnt [30, 31]. In TDT, higher-order moments of  
 376 trait distributions are connected with patterns of variability in the local environment and climate. While efforts to link  
 377 MaxEnt with such environmental variables are ongoing, it may be that constraints based on higher-order moments will  
 378 prove to be of use in the near future in forecasting how biological organisms will change in response to climate forcings.  
 379 Finally, this problem is of interest because even though we start with a different constraint equation than Eq. (1), we  
 380 arrive at the same final result that the most fair distribution is the uniform distribution.

381 To begin we will point out that we will still be using Eqs. (49) and (50) for the specified versions of the probability  
 382 distributions and the partition function. However, we must express the general form of our new constraint equation  
 383 differently via the formula for the standard deviation of a distribution of data as,

$$\sqrt{\sum_{n=1}^{N} p(n) \left[ \left( \sum_{m=1}^{N} f_k(m)p(m) \right) - f_k(n) \right]^2} = \sigma_k(n) \quad (55)$$

384 where  $\sigma_k(n)$  is the standard deviation of our data corresponding to the  $k^{th}$  constraint. Note that, in Eq. (55),  
 385 we have assumed that the mean value of our data,  $\langle f_k(m) \rangle$  can be expressed as  $\sum_{m=1}^{N} f_k(m)p(m)$ , which is the same  
 386 starting expression we have in Eq. (1), and thus may be the reason why our approach is still consistent with the central  
 387 assumptions of MaxEnt.

388 Continuing on as we did in the previous section, we will substitute everything we know into Eq. (55), starting

389 with the fact that  $\sigma_k(n) = \sqrt{2/3}$  for a fair three-sided die (this can be calculated with the definition of the standard  
 390 deviation and a known, equal probability distribution of all outcomes), and that  $K = 1$  for only having one constraint,  
 391  $f_k(n) = f(n) = n, f_k(m) = f(m) = m$ , and  $N = M = 3$ . Doing so gives us,

$$\sqrt{\sum_{n=1}^{n=3} p(n) \left[ \left( \sum_{m=1}^{m=3} m p(m) \right) - n \right]^2} = \sqrt{\frac{2}{3}} \quad (56)$$

392 From here, the steps are quite similar as before, only the algebra is more tedious. We will start with squaring both  
 393 sides of the equation and writing out the summations explicitly to arrive at,

$$\begin{aligned} p(1) [p(1) + 2p(2) + 3p(3) - 1]^2 + \\ p(2) [p(1) + 2p(2) + 3p(3) - 2]^2 + \\ p(3) [p(1) + 2p(2) + 3p(3) - 3]^2 = \frac{2}{3} \end{aligned} \quad (57)$$

394 Substituting  $p(n) = 1/Z e^{-\lambda n} = 1/Z x^n$ , where we have also made the substitution of  $e^{-\lambda} = x$  for simplification, we  
 395 have,

$$\begin{aligned} \frac{x}{Z} \left[ \frac{x}{Z} + 2 \frac{x^2}{Z} + 3 \frac{x^3}{Z} - 1 \right]^2 + \\ \frac{x^2}{Z} \left[ \frac{x}{Z} + 2 \frac{x^2}{Z} + 3 \frac{x^3}{Z} - 2 \right]^2 + \\ \frac{x^3}{Z} \left[ \frac{x}{Z} + 2 \frac{x^2}{Z} + 3 \frac{x^3}{Z} - 3 \right]^2 = \frac{2}{3} \end{aligned} \quad (58)$$

396 Multiplying both sides of the above expression by  $Z^3$  to remove all denominators entirely gives,

$$\begin{aligned} x [x + 2x^2 + 3x^3 - Z]^2 + \\ x^2 [x + 2x^2 + 3x^3 - 2Z]^2 + \\ x^3 [x + 2x^2 + 3x^3 - 3Z]^2 = \frac{2}{3} Z^3 \end{aligned} \quad (59)$$

397 Now comes a lot of distributing and simplifying of terms. Starting with the left-hand-side, and recalling that  
 398  $Z = e^{-\lambda} + e^{-\lambda 2} + e^{-\lambda 3} = x + x^2 + x^3$ , we arrive at,

$$x [x^2 + 2x^3]^2 + x^2 [-x + x^3]^2 + x^3 [-2x - x^2]^2 = \frac{2}{3} [x + x^2 + x^3]^3 \quad (60)$$

399

Expanding both sides of the expression gives,

$$x^4 + 5x^5 + 6x^6 + 5x^7 + x^8 = \frac{2}{3}x^3 + 2x^4 + 4x^5 + \frac{14}{3}x^6 + 4x^7 + 2x^8 + \frac{2}{3}x^9 \quad (61)$$

400

Collecting like terms, we have:

$$\frac{2}{3}x^3 + x^4 - x^5 - \frac{4}{3}x^6 - x^7 + x^8 + \frac{2}{3}x^9 = 0 \quad (62)$$

401

Assuming that  $x = 0$  is not a solution that we want (recall the discussion earlier regarding this possible solution) we can divide by  $x^3$  from both sides to yield,

$$\frac{2}{3} + x - x^2 - \frac{4}{3}x^3 - x^4 + x^5 + \frac{2}{3}x^6 = 0 \quad (63)$$

403

We used Wolfram Alpha online (<https://www.wolframalpha.com/>) to perform the factoring, and find that the above expression can be expressed as,

$$\frac{1}{3}(x - 1)^2(x + 2)(2x + 1)(x^2 + x + 1) = 0 \quad (64)$$

405

These terms produce the five following solutions of  $x = -2$ ,  $x = -1/2$ ,  $x = 1$ , and  $x = -1/2 \pm i\sqrt{3}/2$ . As discussed before, the solutions for negative values of  $x$  are invalid. The solutions  $x = -1/2 \pm i\sqrt{3}/2$  will result in probabilities with non-real (imaginary) terms, so those will also be disregarded. This leaves us with  $x = 1$ , which generates the same uniform distribution of  $p(n) = 1/3$  as we found in the earlier case for constraining only the mean.

409

### A.3 | The Gaussian/Normal Distribution, or Using $n$ and $n^2$ as "Constraint Functions"

410

Here we will show that when the "constraint functions" are  $f_k(n) = n$  and  $f_k(n) = n^2$ , then the resulting probability distribution is Gaussian, or normal. What it means in an ecological context when we specify these constraint functions may not immediately be clear, but there may be some interesting applications of this derivation, as in reconstructing data when publications specify the variance of a measured distribution, but the underlying data have been lost. With or without a useful applied example, following through with the mathematical derivation lends some insight into the behavior of the equations, so it is useful to proceed for that reason. Without specifying an actual problem, we can not solve explicitly for the Lagrange multipliers. Instead, we will investigate what happens to the mathematical form of the probability distribution in Eq. (3) when substituting our constraint information.

418

As we now have two constraints,  $K = 2$ , and we will actually be summing over  $k$  from  $k = 1$  to  $k = 2$ . Additionally, our  $f_k(n)$  now have two forms, one being  $f_1(n) = n$  and the other being  $f_2(n) = n^2$ . Thus, we could write down the set of constraint equations as,

$$\sum_{n=1}^{n=N} np(n) = \langle n \rangle \quad (65)$$

$$\sum_{n=1}^{n=N} n^2 p(n^2) = \langle n^2 \rangle \quad (66)$$

421 However, for this specific example, we don't actually know the values of  $\langle n \rangle$  or  $\langle n^2 \rangle$ , so we will go straight to the  
 422 probability distribution from here. Substituting our constraint functions into the general definition for the probability  
 423 distribution in Eq. (3), we have,

$$p(n) = \frac{1}{Z} e^{-(\lambda_1 n + \lambda_2 n^2)} \quad (67)$$

424 We can extract more insight from this by re-writing the expression by completing the square. We begin by adding  
 425 and subtracting by  $\lambda_1^2/4\lambda_2^2$ , in the argument of the exponential,

$$p(n) = \frac{1}{Z} e^{-\left(\lambda_1 n + \lambda_2 n^2 + \frac{\lambda_1^2}{4\lambda_2^2} - \frac{\lambda_1^2}{4\lambda_2^2}\right)} \quad (68)$$

426 We can now factor the first three terms in the exponential to arrive at,

$$p(n) = \frac{1}{Z} e^{-\left[\left(\frac{\lambda_1}{2\lambda_2} + \lambda_2 n\right)^2 - \left(\frac{\lambda_1}{2\lambda_2}\right)^2\right]} \quad (69)$$

427 This is beginning to look more like the general form of a Gaussian distribution in  $n$ , aside from the pesky constant  
 428 term in the argument of the exponential  $(\lambda_1/2\lambda_2)^2$ . Fortunately, we can use the rules of exponents to rearrange terms  
 429 and remove the second squared term from the argument in the exponential. We then get,

$$p(n) = \frac{e^{\left(\frac{\lambda_1}{2\lambda_2}\right)^2}}{Z} e^{-\left(\frac{\lambda_1}{2\lambda_2} + \lambda_2 n\right)^2} \quad (70)$$

430 Factoring the  $\lambda_2$  term in front of  $n$  gives,

$$p(n) = \frac{e^{\left(\frac{\lambda_1}{2\lambda_2}\right)^2}}{Z} e^{-\frac{\left(n + \frac{\lambda_1}{2\lambda_2}\right)^2}{1/\lambda_2^2}} \quad (71)$$

431 Now we have arrived at an expression that is more easily recognizable as the general form for a Gaussian distribution  
 432 in  $n$ . Comparing this expression to a general Gaussian distribution,

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\sigma^2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (72)$$

433 we can begin to infer what role the Lagrange multipliers have in this scenario. Noting that the term  $\lambda_1/2\lambda_2^2$  sits in the  
 434 place of the distribution mean  $\mu$ , we can conclude that the distribution mean is actually given by the specific combination  
 435 of the Lagrange multipliers of  $\lambda_1/2\lambda_2^2$ . Similarly, noting that the term  $1/\lambda_2^2$  sits in the place of  $2\sigma^2$ , we can conclude  
 436 specifically that the standard deviation of the distribution is given by  $\sigma = 1/\sqrt{2}\lambda_2$ . Matching the overall prefactors tells  
 437 us that the combination of the partition function and our left over constant from completing the square combine to  
 438 normalize the distribution.

#### 439 A.4 | The Log-Normal Distribution, constraining $\log(n)$ and $\log^2(n)$

440 Everything, or more immediately the results, from the above section can be copied and pasted here but replacing every  
 441 instance of  $n$  with  $\log n$  and  $n^2$  with  $\log^2 n$  to give us a log-normal distribution.

## 442 B | APPROXIMATIONS IN THE ORIGINAL VERSION OF METE

443 Two common approximations that are made to simplify the above formulas and provide for limited analytic solutions.  
 444 With the R package "meteR" [28], it is no longer necessary to make approximations to simplify the core equations of  
 445 METE. However, in the original presentation of the theory, a number of important assumptions are used. We derive  
 446 them here because they appear in the Harte (2011) [10], as well as various earlier publications exploring the predictions  
 447 of METE. The steps and assumptions that go into deriving these approximations are worth testing and exploring, as they  
 448 are likely sources of measurable error.

### 449 B.1 | Approximation 1: $\sum e^{-n\beta} \approx 1/\beta$

450 The first approximation used throughout Harte (2011) is presented on pages (149-150) simplifies the relationships  
 451 between the Lagrange multipliers.

$$\sum e^{-n\beta} \approx 1/\beta \quad (73)$$

452 The series  $\sum e^{-n\beta}$  is geometric in the variable  $e^{-\beta}$ . So, as long as  $e^{-\beta} \neq 0$ , then the series can be expressed exactly  
 453 as,

$$\sum_{n=1}^N e^{-n\beta} = \frac{e^{-\beta} - e^{-\beta(N+1)}}{1 - e^{-\beta}} \quad (74)$$

454 Assuming that  $\beta N \gg 1$ , then the second term in the numerator  $e^{-\beta(N+1)} \approx 0$ , giving,

$$\sum_{n=1}^N e^{-n\beta} \approx \frac{e^{-\beta}}{1 - e^{-\beta}} \quad (75)$$

455 Furthermore, assuming that  $\beta \ll 1$ , then  $e^{-\beta} \approx 1 - \beta$ , and we have,

$$\sum_{n=1}^N e^{-n\beta} \approx \frac{1 - \beta}{\beta} \approx 1/\beta \quad (76)$$

456 In deriving the above expression, we made the simultaneous assumptions that  $\beta \ll 1$  and  $N\beta \gg 1$ . This can be  
 457 expressed in the single expression as  $1/N \ll \beta \ll 1$ . There are two steps at which these assumptions are employed,  
 458 thus we could include higher order terms left over from their approximations to get a sense of the resultant order of  
 459 error associated with their use. It should be clear that this approximation breaks down for small  $N$ . "Small  $N$ " will be  
 460 defined in relation to  $\beta$ ; for example, if  $N = 10$ , then  $\beta$  must take on values greater (or much greater) than 0.1. Because  $\beta$   
 461 is defined to be  $\lambda_1 + \lambda_2$ , we can reference Figure 2 to see how likely a possibility this is for all combinations of  $\lambda_1$  and  
 462  $\lambda_2$ . For single data points, this assumption will always break down, but for large data sets  $1/N \ll \beta$  may indeed be a  
 463 valid assumption. However, it should also be clear that  $\beta \ll 1$  will not hold over most of the parameter space that  
 464 is possible for ecosystems (see Figure 2). That said, for all of the empirical data sets that have examined  $\lambda_1$  tends to  
 465 be small (between 0.001 and 0.1), and  $\beta < 1$ , which may be sufficient for these approximations to hold, as advertised,  
 466 approximately.

467 The size of beta will predominately be influenced by  $\lambda_1$  in all cases, and it may be the case that the parameter space  
 468 is more densely populated in the region where  $\lambda_1$  is small (as we begin to see in Figure 2)

## 469 B.2 | Approximation 2: $\sum e^{-n\beta}/n \approx \log(1/\beta)$

470 The series  $\sum_{n=1}^N e^{-n\beta}/n$  is a truncated series expansion for  $\log(1 - x)$ . Specifically, for  $|x| < 1$ ,

$$\log(1 - x) = - \sum_{n=1}^{\infty} \frac{x^n}{n} \quad (77)$$

471 Replacing  $x$  with  $e^{-\beta}$  and breaking the series expansion into two summations, we have,

$$\log(1 - e^{-\beta}) = - \sum_{n=1}^N \frac{e^{-n\beta}}{n} - \sum_{n=N+1}^{\infty} \frac{e^{-n\beta}}{n} \quad (78)$$

472 Note that this expansion requires  $|e^{-\beta}| < 1$ , which is potentially in conflict with the assumption that  $\beta \ll 1$  (an  
 473 assumption used later in this approximation). This conflict is due to the case that the smaller the value  $\beta$  has, the closer  
 474 that  $e^{\beta}$  is to one. Solving the above expression for the partial series expansion, we have

$$\sum_{n=1}^N \frac{e^{-n\beta}}{n} = - \log(1 - e^{-\beta}) + \mathcal{O}\left(\frac{e^{-\beta(N+1)}}{N+1}\right) \quad (79)$$

475 Here, the " $\mathcal{O}$ " represents "order of magnitude of error," which allows us to quantify the error in the truncation of the  
 476 series expansion. From our previous Approximation 1, we are simultaneously assuming that  $N\beta \gg 1$ , thus we will drop  
 477 the  $\mathcal{O}(e^{-\beta(N+1)}/(N+1)$  term to arrive at,

$$\sum_{n=1}^N \frac{e^{-n\beta}}{n} \approx - \log(1 - e^{-\beta}) \quad (80)$$

478 Now, expressing  $e^{-\beta}$  as a series we have,

$$\sum_{n=1}^N \frac{e^{-n\beta}}{n} \approx -\log\left(1 - \sum_{m=1}^{\infty} \frac{(-\beta)^m}{m!}\right) \quad (81)$$

479 Recalling again the assumption that  $\beta \ll 1$ , we can truncate the series expansion of  $e^{-\beta}$  to  $e^{-\beta} \approx 1 + \beta$ , where we have  
480 dropped all terms involving higher powers of  $\beta$ . Doing so, we have

$$\sum_{n=1}^N \frac{e^{-n\beta}}{n} \approx -\log(\beta) \quad (82)$$

481 But,  $-\log(\beta) = 0 - \log(\beta) = \log(1) - \log(\beta) = \log(1/\beta)$ , thus,

$$\sum_{n=1}^N \frac{e^{-n\beta}}{n} \approx \log(1/\beta) \quad (83)$$

482 Because we have used the assumption that  $\beta \ll 1$  to derive this result, and this assumption may not hold in many  
483 cases (see Appendix B.1), both approximations may introduce substantial sources of error in METE's predictions in  
484 some cases, but may otherwise be useful in cases where the user wants to estimate the relative sizes of  $\lambda_1$  and  $\lambda_2$ .  
485 We therefore suggest that calculations involving the Lagrange multipliers be done numerically, without the use of  
486 simplifying approximations, with meteR [28] or a similar software.