Information content and maximum entropy of compartmental systems in equilibrium

Holger Metzler · **Carlos A. Sierra**

Abstract Although compartmental dynamical systems are used in many different areas of science, model selection based on the maximum entropy principle (MaxEnt) is challenging because a lack of methods for quantifying the entropy for this type of systems. Here, we take advantage of the interpretation of compartmental systems as continuous-time Markov chains to obtain entropy measures that quantify model information content. In particular we quantify the uncertainty of a single particle's path as it travels through the system as described by path entropy and entropy rates. Path entropy measures the uncertainty of the entire path of a traveling particle from its entry into the system until its exit, whereas entropy rates measure the average uncertainty of the instantaneous future of a particle while it is in the system. We derive explicit formulas for these two types of entropy for compartmental systems in

Holger Metzler

Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany. Now at: Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Ulls väg 16, 756 51 Uppsala, Sweden, E-mail: holger.metzler@slu.de, ORCiD: 0000-0002-8239-1601

Carlos A. Sierra

Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany, E-mail: csierra@bgc-jena.mpg.de
Department of Ecology, Swedish University of Agricultural Sciences, Ulls väg 16, 756 51 Uppsala, Sweden, ORCiD: 0000-0003-0009-4169

equilibrium based on Shannon information entropy and show how they can be used to solve equifinality problems in the process of model selection by means of MaxEnt.

Keywords Information entropy · Compartmental systems · Equifinality · Model identification · Reservoir models

Mathematics Subject Classification (2000) 34A30 · 60J28 · 60K20 · 92B05

Statements and Declarations

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

1 Introduction

For many modeling applications, it is of interest to quantify the complexity of the system of differential equations used to represent natural phenomena. In principle, we are interested in selecting models that are parsimonious; i.e. with the least degree of complexity for explaining certain patterns in nature. The concept of entropy has been commonly used to characterize complexity or information content. Classical entropy measures for dynamical systems characterize the rate of increase in dynamical complexity as the system evolves over time. These metrics have been used extensively to characterize chaotic behavior in complex nonlinear systems, but they give trivial results for a large range of models used in geosciences and biology.

In a large variety of scientific fields models are based on the principle of mass conservation. In many cases such models are nonnegative dynamical systems that can be described by first-order systems of ordinary differential equations (ODEs) with strong structural constraints. Such systems are called compartmental systems Anderson 1983 Walter and Contreras 1999 Haddad et al. 2010.

Compartmental systems can be evaluated using diagnostic metrics that predict system-level behavior and allow comparisons of systems of very different structures. Age and transit time are two diagnostic metrics of compartmental systems that have been widely studied for systems in and out of equilibrium (Eriksson 1971 Bolin and Rodhe 1973 Rasmussen et al. 2016 Sierra et al. 2016 Metzler and Sierra 2018 Metzler et al. 2018). They help compare behavior and quality of different models. Nevertheless, structurally very different models might show very similar ages and transit times and might represent equally well a given measurement. If we are in the position to choose among such models, which is the one to select? This equifinality problem can be resolved by the maximum entropy principle (MaxEnt) (Jaynes 1957a;b), a generic procedure to draw unbiased inferences from measurement or stochastic data (Pressé et al. 2013).

In order to apply MaxEnt to compartmental systems, some appropriate notion of entropy is required to measure the system's uncertainty or information content. Two classical examples in dynamical systems theory are the topological entropy and the Kolmogorov-Sinai/metric entropy. However, open compartmental systems are dissipative and by Pesin's theorem (Pesin 1977) both metric and topological entropy vanish and cannot serve as a measure of uncertainty. Alternatively, we can interpret compartmental systems as weighted directed graphs. Dehmer and Mowshowitz (2011) provide a comprehensive overview of the history of graph entropy measures. Unfortunately, most of such entropy measures are based on the number of vertices, vertex degree, edges, or degree sequence (Trucco 1956). Thus, they concentrate on only the structural information of the graph. There are also graph theoretical measures that take edges and weights into account by using probability schemes. Their drawback is that the underlying meaning of uncertainty becomes difficult to interpret because the assigned probabilities seem somewhat arbitrary (Bonchev and Buck 2005).

To bridge this gap we introduce three entropy measures based on the Shannon information entropy (Shannon and Weaver 1949) of the continuous-time Markov chain that describes the random path of a single particle through the compartmental system (Metzler and Sierra 2018). While the path entropy describes the uncertainty of a single particle's path through the system, the entropy rate per unit time and the entropy rate per jump describe average uncertainties over the course of a particle's journey. Since this is the first step in this direction, throughout this manuscript we focus on compartmental systems in equilibrium.

The manuscript is organized as follows. First we introduce basic notions of information entropy and compartmental systems in equilibrium together with their associated absorbing continuous-time Markov chain describing the random path of one single particle through the system. Based on this Markov chain we then define three entropy quantities for compartmental systems in equilibrium and adopt the MaxEnt theory. Afterwards we present the introduced theory by means of simple generic examples from the field of carbon-cycle modeling exploring the effect of different parameterizations on the three entropy metrics, before we apply MaxEnt to a model identification problem.

2 Background and theoretical derivation of entropy measures

First, we introduce some basic notations and well-known properties of Shannon information entropy of random variables and stochastic processes. Then, we present compartmental systems as a means to model material-cycle systems that obey the law of mass balance. We then consider such systems from a single-particle point of view and define the path of a single particle through the system along with its visited compartments, sojourn times, occupation times, and transit time. Based on these basic structures of a path, we compute three different types of entropy. For a bet-

ter understanding, we provide a summary of the desirable relations among the three different types of entropy:

(1) As a particle travels through a system of interconnected compartments, it jumps a certain number of times to the next compartment until it finally jumps out of the system. Between two jumps, the particle resides in some compartment. The "path entropy" measures the entire uncertainty about the particles travel through the system, including both the sequence of visited compartments and the respective times spent there.

(2) The entire travel of the particle takes a certain time. In each unit time interval before the particle leaves the system, uncertainties exist whether the particle jumps, where it jumps, and even how often it jumps. The mean of these uncertainties over the mean length of the travel interval is measured by the "entropy rate per unit time".

(3) Each jump comes with uncertainties about which compartment will be next and how long will the particle stay there. The "entropy rate per jump" measures the average of these uncertainties with respect to the mean number of jumps until the particle's exit from the system.

2.1 Short summary of Shannon information entropy

We provide a short introduction of basic concepts of information entropy. For a more detailed introduction along the lines of Cover and Thomas (2006) see Appendix A.

Let Y be a real-valued discrete (continuous) random variable and call p its probability mass function (probability density function). Then

$$\mathbb{H}(Y) := -\mathbb{E}\left[\log p(Y)\right] \tag{1}$$

is called the "Shannon information entropy" ("differential entropy") of Y. Most of the time we just say "entropy" and the precise meaning can be derived from the context.

The entropy's unit depends on the logarithmic base. For base 2 the unit is "bits" and for the natural logarithm with base e the unit is "nats". Throughout this manuscript we use the latter if not stated otherwise.

The entropy $\mathbb{H}(Y)$ of a random variable Y has two intertwined interpretations. On the one hand, it is a measure of uncertainty, that is, a measure of how difficult it is to predict the outcome of a realization of Y. On the other hand, $\mathbb{H}(Y)$ is also a measure of the information content of Y, that is, a measure of how much information we gain once we learn about the outcome of a realization of Y. It is important to note that, even though their definitions and information theoretical interpretations are quite similar, the Shannon- and the differential entropy have one main difference. The Shannon entropy is always nonnegative, whereas the differential entropy can have negative values. While the Shannon entropy is an absolute measure of information and makes sense in its own right, the differential entropy is not an absolute information measure, is not scale-invariant, and makes sense only in comparison with the differential entropy of another random variable.

Panel (a) of Fig. 1 depicts the Shannon entropy with logarithmic base 2 of a Bernoulli random variable Y with $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p \in [0, 1]$ representing a coin toss with probability of heads equal to p. The closer p is to $1/2$ the more difficult it is to predict the outcome, and for an unbiased coin with $p = 1/2$ we have no information about the outcome whatsoever and the Shannon entropy

$$\mathbb{H}(Y) = -p \log p - (1 - p) \log(1 - p) \tag{2}$$

is maximized. Panel (b) of Fig. 1 shows the differential entropy of an exponentially distributed random variable $Y \sim \mathrm{Exp}(\lambda)$ with rate parameter $\lambda > 0$, probability density function $f(y) = \lambda\, e^{-\lambda y}$ for $y \geq 0$, and $\mathbb{E}[Y] = \lambda^{-1}$. We can imagine it to represent the duration of stay of a particle in a well-mixed compartment in an equilibrium compartmental system, where λ is the total outflow rate from the compartment. The

higher the outflow rate is, the more likely an early exit of the particle is, and the easier

it is to predict its moment of exit. Hence, the differential entropy

$$\mathbb{H}(Y) = 1 - \log \lambda \tag{3}$$

decreases with increasing λ.
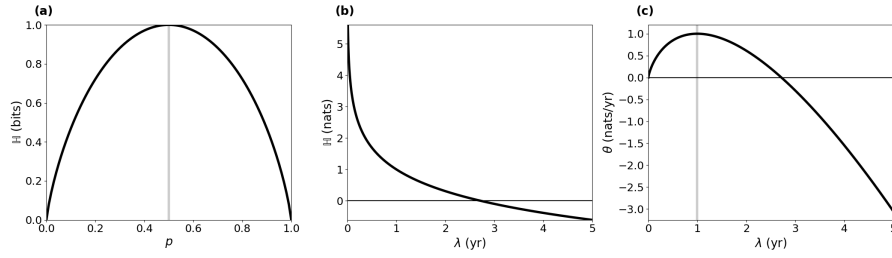
Fig. 1 Shannon entropy of a Bernoulli distribution (a), differential entropy of an exponential distribution (b), and entropy rate of a Poisson process (c). Vertical gray lines indicate the parameter values leading to the highest entropy

We can further consider the "joint entropy" $\mathbb{H}(Y_1, Y_2)$ of two random variables Y_1

and Y_2, which can be described as

$$\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_1) + H(Y_2 \mid Y_1), \tag{4}$$

where the "conditional entropy" $H(Y_2 \mid Y_1)$ describes the uncertainty of Y_2 under the

condition that Y_1 is known. Note that the joint entropy is symmetric, that is, $\mathbb{H}(Y_1, Y_2) =$

$\mathbb{H}(Y_2, Y_1)$ and

$$\mathbb{H}(Y_1, Y_2) \leq \mathbb{H}(Y_1) + \mathbb{H}(Y_2) \tag{5}$$

with equality if Y_1 and Y_2 are independent. Consequently,

$$\mathbb{H}(Y_2 \mid Y_1) \leq \mathbb{H}(Y_2) \tag{6}$$

with equality in case of independence.

The uncertainty of a stochastic process Z can be measured by its "entropy rate" $\theta(Z)$, which describes the time density of the average information in the process. For instance, let $Z \sim \text{Poi}(\lambda)$ be a Poisson process with intensity rate $\lambda > 0$ describing the moments of occurrence of certain events. The interarrival times of Z or the times between events are $\text{Exp}(\lambda)$-distributed, such that in the long run on average the time span between events has length λ^{-1}. The entropy of the interarrival times is given by $\mathbb{H}(\text{Exp}(\lambda)) = 1 - \log \lambda$, and averaging it over the mean interarrival time gives the entropy rate of the Poisson process Z (Gaspard and Wang 1993, Sect. 3.3), that is,

$$\theta(Z) = \theta(\text{Poi}(\lambda)) = \lambda \left(1 - \log \lambda\right). \tag{7}$$

This entropy rate increases with $\lambda \in [0, 1]$, reaches its maximum at 1, and then it decreases (Fig. 1, panel c). This behavior is independent of the unit of λ, because it is based on the differential entropy of the exponential distribution and hence not scale-invariant. Consequently, it is not an absolute measure of information content, but only useful in comparison to the entropy rates of other stochastic processes.

2.2 Compartmental systems in equilibrium

Mass-balanced flow of material into a system, within the system and out of the system that consists of several compartments can be modeled by so-called compartmental systems (Anderson 1983 Jacquez and Simon 1993). Compartments are always well-mixed and usually also called "pools" or "boxes". One way to describe compartmental systems is by the d-dimensional linear ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{x}(t) = \mathrm{B}\,\mathbf{x}(t) + \mathbf{u}, \quad t > 0, \tag{8}$$

with some initial condition $\mathbf{x}(0) = \mathbf{x}^0 \in \mathbb{R}^d$. The nonnegative vector $\mathbf{x}(t)$ describes the amount of material in the different compartments at time t, the nonnegative vector $\mathbf{u}$

is the vector of external inputs to the compartments, and the compartmental matrix $B \in \mathbb{R}^{d \times d}$ describes the flux rates between the compartments and out of the system. The nonnegative off-diagonal value B_{ij} is the flux rate from compartment j to compartment i, the absolute value of the negative diagonal value B_{jj} is the total rate of fluxes out of compartment j, and the nonnegative value $z_j = -\sum_{i=1}^{d} B_{ij}$ is the rate of the flux from compartment j out of the system. By requiring B to be invertible we ensure that the system is "open", that is, all material that enters the system will eventually also leave it. Throughout this manuscript, we consider the open compartmental system (8) to have reached its unique steady-state or equilibrium compartment vector $\mathbf{x}^* = -B^{-1}\mathbf{u}$. An open compartmental system in equilibrium given by Eq.(8) is fully characterized by $\mathbf{u}$ and B, and we denote it by $M = M(\mathbf{u}, B)$.

2.3 The one-particle perspective

While Eq. (8) describes the movement of bulk material through the system, compartmental systems in equilibrium can also be described probabilistically by considering the random path of a single particle through the system (Metzler and Sierra 2018). If $X_t \in \mathscr{S} := \{1, 2, \ldots, d\}$ denotes the compartment in which the single particle is at time t, and $X_t = d+1$ if the particle has already left the system, then $X := (X_t)_{t \geq 0}$ is an absorbing continuous-time Markov chain (Norris 1997) on $\widetilde{\mathscr{S}} := \mathscr{S} \cup \{d+1\}$. Its initial distribution is given by $\widetilde{\beta} = (\beta_1, \beta_2, \ldots, \beta_d, 0)^T$, where $\beta := \mathbf{u}/\|\mathbf{u}\|$, and $\beta_j = \mathbb{P}(X_0 = j)$ is the probability of the single particle to enter the system through compartment j. The superscript T denotes the transpose of the vector/matrix and $\|\mathbf{u}\| = \sum_{i=1}^{d} |u_i|$ denotes the l_1-norm of the vector $\mathbf{u}$. The state-transition matrix of X is given by

$$Q = \begin{pmatrix} B & \mathbf{0} \\ \mathbf{z}^T & 0 \end{pmatrix}, \tag{9}$$

and thus

$$\mathbb{P}(X_t = i) = (e^{t\mathrm{Q}}\widetilde{\beta})_i = \sum_{j=1}^{d}(e^{t\mathrm{Q}})_{ij}\beta_j, \quad i \in \widetilde{\mathscr{S}}, \tag{10}$$

is the probability of the particle to be in compartment i at time t if $i \in \mathscr{S}$ or that the particle has left the system if $i = d + 1$. Here, $e^{t\mathrm{Q}}$ denotes the matrix exponential. Furthermore,

$$\mathbb{P}(X_t = i | X_s = j) = (e^{(t-s)\mathrm{Q}})_{ij}, \quad s \le t, \quad i,j \in \widetilde{\mathscr{S}}, \tag{11}$$

is the probability that X is in state i at time t given it was in state j at time s. Since the Markov chain X and the compartmental system in equilibrium given by Eq. (8) are equivalent, we can write

$$M = M(\mathbf{u}, \mathrm{B}) = M(X). \tag{12}$$

2.4 The path of a single particle

A particle's path through the system from the moment of entering until the moment of exit can be described as a sequence of (compartment, sojourn-time)-pairs

$$\mathscr{P}(X) := ((Y_1 = X_0, T_1), (Y_2, T_2), \ldots, (Y_{\mathscr{N}-1}, T_{\mathscr{N}-1}), Y_{\mathscr{N}} = d+1), \tag{13}$$

where X is the absorbing Markov chain associated to the particle's journey. The sequence $Y_1, Y_2, \ldots, Y_{\mathscr{N}-1} \in \mathscr{S}$ represents the successively visited compartments along with the associated sojourn times $T_1, T_2, \ldots, T_{\mathscr{N}-1}$, the random variable

$$\mathscr{N} := \inf\{n \in \mathbb{N} : Y_n = d+1\} \tag{14}$$

denotes the first hitting time of the absorbing state $d + 1$ by the "embedded jump chain" $Y := (Y_n)_{n=1,2,\ldots,\mathscr{N}}$ of X (Norris 1997). With $\lambda_j := -Q_{jj}$ the one-step transi-

tion probabilities of Y are given by, for $i, j \in \widetilde{\mathscr{S}}$,

$$P_{ij} := \mathbb{P}(Y_{n+1} = i \,|\, Y_n = j) = \begin{cases} 0, & i = j \text{ or } \lambda_j = 0, \\ Q_{ij}/\lambda_j, & \text{else.} \end{cases} \tag{15}$$

We can also write $\mathrm{P} = (P_{ij})_{i,j \in \widetilde{\mathscr{S}}} = \mathrm{Q}\mathrm{D}^{-1} + \mathrm{I}$, where

$$\mathrm{D} := \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_d, \lambda_{d+1}) \tag{16}$$

is the diagonal matrix with the diagonal entries of Q and I denotes the identity matrix of appropriate dimension. We define the matrix $\mathrm{P_B} := (P_{ij})_{i,j \in \mathscr{S}}$, then $\mathrm{M} := (\mathrm{I} - \mathrm{P_B})^{-1}$ is the "fundamental matrix" of Y, with $\mathrm{I} \in \mathbb{R}^{d \times d}$ denoting the identity matrix. The entry M_{ij} denotes the expected numbers of visits to compartment i given that the particle entered the system through compartment j. Consequently, the expected number of visits to compartment $i \in \mathscr{S}$ is given by

$$N_i = \sum_{j=1}^d M_{ij}\beta_j = (\mathrm{M}\beta)_i = \left[(\mathrm{I} - \mathrm{P_B})^{-1}\beta\right]_i = (\mathrm{D}\mathrm{B}^{-1}\beta)_i = \frac{\lambda_i x_i^*}{\|\mathbf{u}\|} \tag{17}$$

and the total expected number of jumps is given by

$$\mathbb{E}\,[\mathscr{N}] = \sum_{i=1}^d (\mathrm{M}\beta)_i + 1 = \sum_{i=1}^d N_i + 1, \tag{18}$$

where we take into account also the last jump out of the system.

The last jump, $\mathscr{N}$, leads the particle out of the system such that at the moment of this last jump X takes on the value $d + 1$. This last jump happens at the absorption time of the Markov chain X, which is defined as

$$\mathscr{T} := \inf\{t > 0 : X_t = d + 1\}. \tag{19}$$

The absorption time is phase-type distributed (Neuts 1981), $\mathscr{T} \sim \mathrm{PH}(\beta, \mathrm{B})$, with probability density function

$$f_{\mathscr{T}}(t) = \mathbf{z}^T e^{t\,\mathrm{B}} \beta, \quad t \geq 0. \tag{20}$$

It can be shown (Metzler and Sierra 2018, Sect. 3.2) that the mean or expected value of $\mathscr{T}$ equals the turnover time (Sierra et al. 2016) of system (8) in equilibrium and is given by total stocks over total fluxes, that is,

$$\mathbb{E}[\mathscr{T}] = \frac{\|\mathbf{x}^*\|}{\|\mathbf{u}\|}. \tag{21}$$

Furthermore, by construction $\sum_{k=1}^{\mathcal{N}-1} T_k = \mathscr{T}$. If we denote by $\mathbb{1}_{\{A\}}$ the indicator function of the logical expression A, given by

$$\mathbb{1}_{\{A\}} = \begin{cases} 1, & A \text{ is true,} \\ 0, & \text{else,} \end{cases} \tag{22}$$

then $O_j := \sum_{k=1}^{\mathcal{N}-1} \mathbb{1}_{\{Y_k=j\}} T_k$ is the total time that the particle spends in compartment j. This time is called "occupation time" of j and its mean is given by (Metzler and Sierra 2018, Sect. 3.3)

$$\mathbb{E}[O_j] = \frac{x_j^*}{\|\mathbf{u}\|}, \tag{23}$$

which induces $\mathbb{E}[\mathscr{T}] = \sum_{j=1}^{d} \mathbb{E}[O_j]$.

2.5 Path entropy, entropy rate per unit time, entropy rate per jump

The path $\mathscr{P} = \mathscr{P}(X)$ given by Eq. (13) can be interpreted in three different ways. Each of these ways leads to a different interpretation of the path's entropy. First, we can look at $\mathscr{P}$ as the result of bookkeeping of the absorbing continuous-time Markov chain X, where for all times t we note down the pair (X_t, t) of the current compartment

and the current time. Second, we can consider the path as a discrete-time process. In each time step n, we choose randomly a new compartment Y_{n+1} and an associated sojourn time T_{n+1} of the particle in this compartment. Third, we can look at $\mathscr{P}$ as a single random variable with values in the space of all possible paths. Based on the latter interpretation we now derive the path entropy.

We are interested in the uncertainty/information content of the path $\mathscr{P}(X)$ of a single particle. Along the lines of Albert (1962), we construct a space $\wp$ that contains all possible paths that can be taken by a particle that runs through the system until it leaves. Let $\wp_n := (S \times \mathbb{R}_+)^n \times \{d+1\}$ denote the space of paths that visit n compartments/states before ending up in the environmental compartment/absorbing state $d+1$. By $\wp := \bigcup_{n=1}^{\infty} \wp_n$ denote the space of all eventually absorbed paths. Note that, since B is invertible, a path through the system is finite with probability 1. Let l denote the Lebesgue measure on $\mathbb{R}_+$ and c the counting measure on S. Furthermore, let σ_n be the sigma-finite product measure on $\wp_n$. It is defined by $\sigma_n := (c \otimes l)^n \otimes c$. Almost all sample functions of $(X_t)_{t \geq 0}$ can be represented as a point $p \in \wp$ (Doob 1953, Chapter VI). Consequently, we can represent X by a finite-length path $\mathscr{P}(X) = ((Y_1, T_1), (Y_2, T_2), \ldots, (Y_n, T_n), Y_{n+1})$ for some $n \in \mathbb{N}$, where $Y_{n+1} = d+1$.

For each set $W \subseteq \wp$ for which $W \cap \wp_n$ is σ_n-measurable for each $n \in \mathbb{N}$, we define $\sigma^*(W) := \sum_{n=1}^{\infty} \sigma_n(W \cap \wp_n)$. This measure is defined on the σ-field $\mathscr{F}^*$ which is the smallest σ-field containing all sets $W \subseteq \wp$ whose projection on $\mathbb{R}_+^n$ is a Borel set for each $n \in \mathbb{N}$. Let σ be a measure on all sample functions, defined for all subsets W whose intersection with $\wp$ is in $\mathscr{F}^*$. We define it by $\sigma(W) := \sigma^*(W \cap \wp)$.

Let $p = ((x_1, t_1), (x_2, t_2,), \ldots, (x_n, t_n), d+1) \in \wp$ for some $n \in \mathbb{N}$. For $i \neq j$, denote by $N_{ij}(p)$ the total number of path p's one-step transitions from j to i and by $R_j(p)$ the total amount of time spent in j.

Theorem 1 The probability density function of $\mathscr{P} = \mathscr{P}(X)$ with respect to σ is given by

$$f_{\mathscr{P}}(p) = \beta_{x_1} \left(\prod_{j=1}^{d} \prod_{i=1, i \neq j}^{d+1} (Q_{ij})^{N_{ij}(p)} \right) \prod_{j=1}^{d} e^{-\lambda_j R_j(p)},$$

$$p = ((x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n), d+1) \in \wp. \tag{24}$$

Proof Let $x_1, x_2, \ldots, x_n \in S$, $x_{n+1} = d+1$, and $t_1, t_2, \ldots, t_n \in \mathbb{R}_+$. Since

$$\mathbb{P}((Y_1 = x_1, T_1 \leq t_1), \ldots, (Y_n = x_n, T_n \leq t_n), Y_{n+1} = d+1)$$

$$= \mathbb{P}(Y_{n+1} = d+1 \,|\, Y_n = x_n)$$

$$\cdot \prod_{k=2}^{n} \mathbb{P}(Y_k = x_k, T_k \leq t_k \,|\, Y_{k-1} = x_{k-1}) \mathbb{P}(Y_1 = x_k, T_1 \leq t_1)$$

$$= P_{d+1, x_n} \left[\prod_{k=2}^{n} P_{x_k x_{k-1}} \left(1 - e^{-\lambda_{x_k} t_k} \right) \right] \beta_{x_1} \left(1 - e^{-\lambda_{x_1} t_1} \right)$$

$$= \int_{\mathbb{T}_n} \beta_{x_1} \prod_{k=1}^{n} Q_{x_{k+1} x_k} e^{-\lambda_{x_k} \tau_k} \, d\tau_1 d\tau_2 \cdots d\tau_n \tag{25}$$

with $\mathbb{T}_n = \{(\tau_1, \tau_2, \ldots, \tau_n) \in \mathbb{R}_+^n : 0 \leq \tau_1 \leq t_1, 0 \leq \tau_2 \leq t_2, \ldots, 0 \leq \tau_n \leq t_n\}$, the probability density function of $\mathscr{P} = \mathscr{P}(x)$ with respect to σ is given by

$$f_{\mathscr{P}}(p) = \beta_{x_1} \prod_{k=1}^{n} Q_{x_{k+1} x_k} e^{-\lambda_{x_k} t_k},$$

$$p = ((x_1, t_1), (x_2, t_2), \ldots, (x_n, t_n), d+1) \in \wp. \tag{26}$$

The term $Q_{x_{k+1} x_k} = Q_{ij}$ enters exactly $N_{ij}(p)$ times. Furthermore,

$$\prod_{k=1}^{n} e^{-\lambda_{x_k} t_k} = \prod_{k=1}^{n} \prod_{j=1}^{d} \mathbb{1}_{\{x_k = j\}} e^{-\lambda_j t_k} = \prod_{j=1}^{d} e^{-\lambda_j \sum_{k=1}^{n} \mathbb{1}_{\{x_k = j\}} t_k}$$

$$= \prod_{j=1}^{d} e^{-\lambda_j R_j(p)}. \tag{27}$$

We make the according substitutions and the proof is finished.

The entropy of the absorbing continuous-time Markov chain X is equal to its entropy on the random but finite time horizon $[0, \mathcal{T}]$, which in turn equals the entropy of a single particle's path $\mathscr{P}$ through the system.

Theorem 2 The entropy of the absorbing continuous-time Markov chain X is given by

$$
\begin{aligned}
\mathbb{H}(X) = \mathbb{H}(\mathscr{P}) \\
= -\sum_{i=1}^{d} \beta_i \log \beta_i \\
+ \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1,i\neq j}^{d} B_{ij}\left(1 - \log B_{ij}\right) + z_j\left(1 - \log z_j\right) \right].
\end{aligned}
\tag{28}
$$

Proof Let X have the finite path representation

$$
\mathscr{P} = \mathscr{P}(X) = ((Y_1, T_1), (Y_2, T_2), \ldots, (Y_n, T_n), d+1)
\tag{29}
$$

for some $n \in \mathbb{N}$, and denote by $f_{\mathscr{P}}$ its probability density function. Then, by Theorem 1,

$$
-\log f_{\mathscr{P}}(\mathscr{P}) = -\log \beta_{Y_1} - \sum_{j=1}^{d} \sum_{i=1,i\neq j}^{d+1} N_{ij}(\mathscr{P}) \log Q_{ij} + \sum_{j=1}^{d} \lambda_j R_j(\mathscr{P}).
\tag{30}
$$

We compute the expectation and get

$$
\begin{aligned}
\mathbb{H}(X) = \mathbb{H}(\mathscr{P}) = -\mathbb{E}\left[\log f_{\mathscr{P}}(\mathscr{P})\right] \\
= -\mathbb{E}\left[\log \beta_{Y_1}\right] - \sum_{j=1}^{d} \sum_{i=1,i\neq j}^{d+1} \mathbb{E}\left[N_{ij}(\mathscr{P})\right] \log Q_{ij} + \sum_{j=1}^{d} \lambda_j \mathbb{E}\left[R_j(\mathscr{P})\right] \\
= \mathbb{H}(Y_1) + \sum_{j=1}^{d} \lambda_j \mathbb{E}\left[R_j(\mathscr{P})\right] - \sum_{j=1}^{d} \sum_{i=1,i\neq j}^{d+1} \mathbb{E}\left[N_{ij}(\mathscr{P})\right] \log Q_{ij}.
\end{aligned}
\tag{31}
$$

Obviously, $\mathbb{E}\left[R_j(\mathscr{P})\right] = \mathbb{E}\left[O_j\right] = x_j^*/\|\mathbf{u}\|$ is the mean occupation time of compartment $j \in S$ by X. Furthermore, for $i \in \widetilde{S}$ and $j \in S$ such that $i \neq j$, by Eqs. (17)

and (15),

$$\mathbb{E}\left[N_{ij}(\mathscr{P})\right] = \mathbb{E}\left[N_j(\mathscr{P})\right] P_{ij} = \begin{cases} \frac{x_j^*}{\|\mathbf{u}\|} B_{ij}, & i \leq d, \\ \\ \frac{x_j^*}{\|\mathbf{u}\|} z_j, & i = d+1. \end{cases} \tag{32}$$

Together with $\lambda_j = \sum_{i=1, i \neq j}^{d} B_{ij} + z_j$, we obtain

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(Y_1) + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\left(\sum_{i=1, i \neq j}^{d} B_{ij} + z_j \right) \right. \\ &\qquad\qquad \left. - \sum_{i=1, i \neq j}^{d} B_{ij} \log B_{ij} - z_j \log z_j \right] \\ &= -\sum_{i=1}^{d} \beta_i \log \beta_i + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^{d} B_{ij} \left(1 - \log B_{ij} \right) \right. \\ &\qquad\qquad\qquad \left. + z_j \left(1 - \log z_j \right) \right]. \end{aligned} \tag{33}$$

By some simple substitutions and rearrangements, we obtain two representations of $\mathbb{H}(X) = \mathbb{H}(\mathscr{P})$ that are easy to interpret. For simplicity of notation, we define

$$\mathbb{H}(\beta) := -\sum_{i=1}^{d} \beta_i \log \beta_i. \tag{34}$$

Proposition 1 The entropy of the absorbing continuous-time Markov chain X is also given by

$$\mathbb{H}(X) = \mathbb{H}(\beta) + \sum_{j=1}^{d} \mathbb{E}\left[O_j\right] \left(\sum_{i=1, i \neq j}^{d} \theta(\mathrm{Poi}(B_{ij})) + \theta(\mathrm{Poi}(z_j)) \right) \tag{35}$$

and

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\beta) \\ &\quad + \sum_{j=1}^{d} \mathbb{E}\left[N_j\right] \left(\mathbb{H}(\mathrm{Exp}(\lambda_j)) + \mathbb{H}(P_{1,j}, P_{2,j}, \ldots, P_{d,j}, P_{d+1,j}) \right), \end{aligned} \tag{36}$$

which can be rewritten as

$$\mathbb{H}(X) = \mathbb{H}(\beta) + \sum_{j=1}^{d} \mathbb{E}[N_j] \, \mathbb{H}(P_{1,j}, P_{2,j}, \ldots, P_{d,j}, P_{d+1,j}) \tag{37}$$

$$+ \sum_{j=1}^{d} \mathbb{E}[N_j] \, \mathbb{H}(\mathrm{Exp}(\lambda_j)). \tag{38}$$

Proof By virtue of Eq. (35) we replace $x_j^*/\|\mathbf{u}\|$ by $\mathbb{E}[O_j]$ in Eq. (28) and take into account that the entropy rate of a Poisson process with intensity rate λ equals $\lambda(1 - \log\lambda)$ to prove Eq. (35). To prove Eq. (36) we use Eq. (17) to replace $x_j^*/\|\mathbf{u}\|$ in Eq. (28) by $\mathbb{E}[N_j]/\lambda_j$ and obtain

$$\begin{aligned}
\mathbb{H}(X) = &- \sum_{i=1}^{d} \beta_i \log \beta_i + \sum_{j=1}^{d} \mathbb{E}[N_j] \, (1 - \log\lambda_j) \\
&+ \sum_{j=1}^{d} \mathbb{E}[N_j] \left(- \sum_{i=1, i\neq j}^{d} \frac{B_{ij}}{\lambda_j} \log \frac{B_{ij}}{\lambda_j} - \frac{z_j}{\lambda_j} \log \frac{z_j}{\lambda_j} \right).
\end{aligned} \tag{39}$$

Here, $(1 - \log\lambda_j)$ is the entropy of an exponential random variable with rate parameter λ_j. Using definition (15) of P_{ij} we replace B_{ij}/λ_j by P_{ij} for $i \in \mathscr{S}$ and z_j/λ_j by $P_{d+1,j}$ and finish the proof.

We now define the "path entropy" of the compartmental system M in equilibrium, given by Eq. (8), as the path entropy of its associated continuous-time Markov chain X, that is

$$\mathbb{H}(M) := \mathbb{H}(X) = \mathbb{H}(\mathscr{P}(X)). \tag{40}$$

For a one-dimensional compartmental system M_λ in equilibrium with rate $\lambda > 0$ and positive external input given by

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = -\lambda \, x(t) + u, \quad t > 0, \tag{41}$$

the entropy of the initial distribution vanishes, and we obtain

$$\mathbb{H}(M_\lambda) = \frac{x^*}{u} \lambda (1 - \log \lambda) = \frac{1}{\lambda} \lambda (1 - \log \lambda) = 1 - \log \lambda, \qquad (42)$$

which equals the differential entropy $1 - \log \lambda$ of the exponentially distributed mean transit time $\mathscr{T}_\lambda \sim \text{Exp}(\lambda)$, reflecting that the only uncertainty of the particle's path in a one-pool system is the time of the particle's exit. The exponential distribution with rate parameter λ is the distribution of the interarrival time of a Poisson process wit intensity rate λ. Hence, we can interpret $\mathbb{H}(M_\lambda) = \lambda^{-1} \lambda (1 - \log \lambda)$ as the instantaneous Poisson entropy rate $\lambda (1 - \log \lambda)$ multiplied with the expected duration $\mathbb{E}[\mathscr{T}] = \lambda^{-1}$ of the particle's stay in the system.

Migrating to a d-dimensional system, we can interpret $\mathbb{H}(M)$ as the entropy of a continuous-time process in the light of Eq. (35) and as the entropy of a discrete-time process in the light of Eq. (36). In both interpretations, the first term $\mathbb{H}(\beta) = \mathbb{H}(X_0) = \mathbb{H}(Y_1)$ represents the uncertainty of the first pool through which the particle enters the system. In the continuous-time interpretation, the uncertainty of the subsequent travel is the weighted average of the superposition of d Poisson processes describing the instantaneous uncertainty of possible jumps of the particle inside the system, $\theta(\text{Poi}(B_{ij}))$, and out of the system, $\theta(\text{Poi}(z_j))$, where the weights are the expected occupation times of the different compartments $j \in \mathscr{S}$. In the discrete-time interpretation, the subsequent travel's uncertainty is the average of uncertainties associated to each pool, weighted by the number of visits to the respective pools. The uncertainty associated to each pool comprises the uncertainty of the length of the stay in the pool, $\mathbb{H}(\text{Exp}(\lambda_j))$, and the uncertainty of where to jump afterwards, $\mathbb{H}(\{P_{ij} : i \in \widetilde{\mathscr{S}}, j \in \mathscr{S}, i \neq j\})$. Hence, in the light of Eq. (36), we can separate the path entropy into a discrete part associated to jump uncertainty given by Eq. (37) and a continuous part associated to sojourn time uncertainty given by Eq. (38).

The two interpretations of the path entropy $\mathbb{H}(M)$ (as a continuous-time or discrete-time process) motivate two different entropy rates as described earlier. The "entropy rate per unit time" is given by

$$\theta(M) = \frac{\mathbb{H}(M)}{\mathbb{E}\left[\mathscr{T}\right]} \tag{43}$$

and the "entropy rate per jump" by

$$\theta_J(M) = \frac{\mathbb{H}(M)}{\mathbb{E}\left[\mathscr{N}\right]}. \tag{44}$$

While the path entropy measures the uncertainty of the entire path, entropy rates measure the average uncertainty of the instantaneous future of a particle while it is in the system: for the entropy rate per unit time the uncertainty entailed by the infinitesimal future, and for the entropy rate per jump the uncertainty entailed by the next jump.

2.6 The maximum entropy principle (MaxEnt)

MaxEnt arose in statistical mechanics as a variational principle to predict the equilibrium states of thermal systems and later was applied to matters of information and as a general procedure to draw inferences based on self-consistency requirements (Pressé et al. 2013). Its relationship to information theory and stochastics was established by Jaynes (1957a;b). The general idea is to identify the most uninformed probability distribution to represent some given data in the sense that the maximum entropy distribution, constrained to given data, uses the information provided by the data only and nothing else. This approach ensures that no additional subjective information creeps into the distribution. The goal of this section is to transfer MaxEnt to compartmental systems in order to identify the compartmental system that represents

our state of knowledge best in different situations, and at the same time get a better understanding of the introduced entropy measures.

Example 1 Consider the set $\mathscr{M}_1$ of equilibrium compartmental systems (8) with a predefined nonzero input vector $\mathbf{u}$, a predefined mean transit time $\mathbb{E}[\mathscr{T}]$, and an unknown steady-state vector $\mathbf{x}^*$ comprising nonzero components. We are interested in the most unbiased compartmental system that reflects our state of information, where maximum unbiasedness is achieved by identifying $M_1^* \in \mathscr{M}_1$ such that the entropy rate per unit time $\theta(M_1^*)$, or equivalently the path entropy $\mathbb{H}(\mathscr{P}(M_1^*))$, is maximized. We can show (see Proposition B.1) that the compartmental system $M_1^* = M(\mathbf{u}, B)$ with

$$
B = \begin{pmatrix} -\lambda & 1 & \cdots & 1 \\ 1 & -\lambda & 1 \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & -\lambda \end{pmatrix}, \tag{45}
$$

where $\lambda = d - 1 + 1/\mathbb{E}[\mathscr{T}]$, is the maximum entropy model in $\mathscr{M}_1$. In the special case $d = 1$ for a one-dimensional compartmental system, we obtain $B = -1/\mathbb{E}[\mathscr{T}]$. Since in this case $\mathscr{T} \sim \mathrm{Exp}(-B)$, we see that the exponential distribution is the maximum entropy distribution in the class of all nonnegative continuous probability distributions with fixed expected value. This special case is very well known (Cover and Thomas 2006, Example 12.2.5).

Example 2 Let us consider the sub§class $\mathscr{M}_2 \subseteq \mathscr{M}_1$ of compartmental models from the previous example with the additional restriction of a predefined positive steady-state vector $\mathbf{x}^*$. Then the compartmental system $M_2^* = M(\mathbf{u}, B)$ with

$$
B_{ij} = \begin{cases} \sqrt{\dfrac{x_i^*}{x_j^*}}, & i \neq j, \\ -\displaystyle\sum_{k=1, k \neq j}^{d} \sqrt{\dfrac{x_k^*}{x_j^*}} - \dfrac{1}{\sqrt{x_j^*}}, & i = j, \end{cases} \tag{46}
$$

365 is the maximum entropy model in $\mathscr{M}_2$ (see Proposition B.2).

366 2.7 Structural model identification via MaxEnt

367 Suppose we observe a natural system and conduct measurements from which we try

368 to construct a linear autonomous compartmental model in equilibrium that represents

369 the observed natural system as well as possible. The first question that arises is the

370 one for the number of compartments the model should ideally have. MaxEnt cannot

371 be helpful here because by adding more and more compartments we can theoretically

372 increase the entropy of the model indefinitely. Consequently, the problem of finding

373 the right dimension of system (8) has to be solved by other means. One way to do this

374 is to analyze an impulse response function of the system and its Laplace transform,

375 that is the transfer function of the system, and identify the most dominating frequen-

376 cies. The impulse response or the transfer function might be possible to obtain by

377 tracer experiments (Anderson 1983 Walter 1986).

378   In Anderson (1983, Chapter 16) the "structural identification problem" of linear

379 autonomous systems is described as follows. Suppose we are interested in determin-

380 ing a d-dimensional system of form (8). We are interested in sending an impulse into

381 the system at time $t = 0$ and analyzing its further behavior. To that end, we rewrite

382 the system to

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{x}(t) = B\,\mathbf{x}(t) + A\,\mathbf{u}, \quad t \geq 0,$$

$$\mathbf{x}(0) = \mathbf{0} \in \mathbb{R}^d, \tag{47}$$

$$\mathbf{y}(t) = C\,\mathbf{x}(t), \quad t \geq 0.$$

Note that the roles of A and B are interchanged here with respect to Anderson (1983).

In a typical tracer experiment, we choose an input vector $\mathbf{u}$ and the "input distribution

matrix" A, which defines how the input vector enters the system. Then we decide

which compartments we can observe to determine the "output connection matrix" C.

The experiment is now to inject an impulse into the system and to record the output

function $\mathbf{y}(t) = C\mathbf{x}(t)$. Bellman and Åström (1970) pointed out that the input-output relation is given by

$$\mathbf{y}(t) = C\mathbf{x}(t) = C \int_0^t e^{(t-\tau)B} A\mathbf{u}(\tau)\,d\tau$$

$$= \left[C e^{tB} A\right] * \mathbf{u}(t),$$

where $*$ is the convolution operator. The model parameters enter the input-output relation only in the matrix-valued "impulse response function"

$$\Psi(t) := C e^{tB} A, \quad t \ge 0, \tag{48}$$

or in the "transfer function"

$$\widehat{\Psi}(s) := C(sI - B)^{-1} A, \quad s \ge 0, \tag{49}$$

which is the Laplace transform matrix of Ψ. Consequently, all identifiable parameters of A, B, and C must be identified through Ψ or $\widehat{\Psi}$. Difficulties arise because the entries of the matrices Ψ and $\widehat{\Psi}$ are usually nonlinear expressions of the elements of A, B, and C. We call system (47) "identifiable" if this nonlinear system of equations has a unique solution (A, B, C) for given Ψ or $\widehat{\Psi}$. Otherwise the system is called "non-identifiable". Usually, the matrices A and C are already know from the experiment's setup. What remains is to identify the compartmental matrix B, and this can be done by MaxEnt.

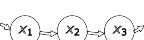## 3 Application to particular systems

First, we apply the presented theory to some equilibrium compartmental models with very simple structure in order to get some grasp on the new entropy concept. Then we compute entropy quantities for two carbon-cycle models in dependence on en-

vironmental and biochemical parameters. Afterwards, we apply MaxEnt to solve an equifinality problem in model selection.

3.1 Simple examples

From Table 1 we can see that depending on the connections between compartments smaller systems can have greater path entropy and entropy rates than bigger systems, even though systems with more compartments can theoretically reach higher entropy. Furthermore, we see from the depicted examples that the system with the highest path entropy does neither have the highest entropy rate per unit time nor per jump. Adding connections to a system, one would expect higher path entropy, but the path entropy might actually decrease because the new connections potentially provide a faster way out the system.

Structure	$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t)$	θ_J	$\mathbb{E}[\mathcal{N}]$	θ	$\mathbb{E}[\mathcal{T}]$	$\mathbb{H}(\mathcal{P})$
x_1	$-\lambda\,x + 1$	$0.5\,(1 - \log\lambda)$	2.00	$\lambda\,(1 - \log\lambda)$	$1/\lambda$	$1 - \log$
$x_1 \to x_2$	$\begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	0.67	3.00	1.00	2.00	2.00
$x_1\ x_2$	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	0.85	2.00	1.69	1.00	1.69
$x_1 \rightleftarrows x_2$	$\begin{pmatrix} -1 & 0.5 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	1.08	5.00	1.35	_4.00_	_5.39_
$x_1 \rightleftarrows x_2$	$\begin{pmatrix} -1 & 0.5 \\ 0.5 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	_1.36_	3.00	2.04	2.00	4.08
$x_1 \to x_2 \to x_3$	$\begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	0.75	4.00	1.00	3.00	3.00
$x_1\ x_2\ x_3$	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	1.05	2.00	_2.10_	1.00	2.10

Table 1 Overview of different entropy measures of simple models with different structures. Underlined numbers are the highest values per column

3.2 A linear autonomous global carbon-cycle model

We consider the global carbon-cycle model introduced by Emanuel et al. (1981)
(Fig. 2). The model comprises five compartments: non-woody tree parts $x_1 = 37\,\mathrm{PgC}$,
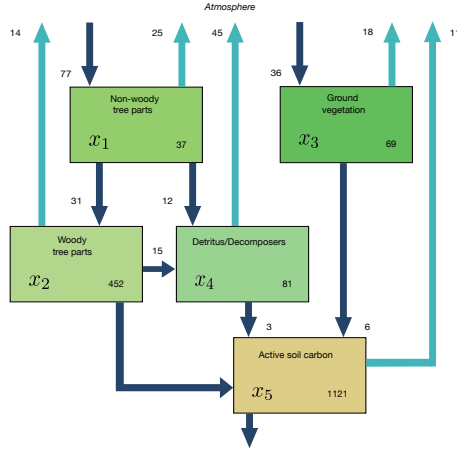
Fig. 2 Schematic of the linear autonomous global carbon cycle model in steady state introduced by
Emanuel et al. (1981)

woody tree parts $x_2 = 452\,\mathrm{PgC}$, ground vegetation $x_3 = 69\,\mathrm{PgC}$, detritus/decomposers
$x_4 = 81\,\mathrm{PgC}$, and active soil carbon $x_5 = 1,121\,\mathrm{PgC}$. We introduce an environmental
rate modifier ξ which controls the speed of the system. This parameter could poten-
tially increase and speed up the system with increasing global surface temperature.
For a given ξ the equilibrium model $M_\xi = M(\mathbf{u}, \mathrm{B}_\xi)$ is given by

$$\mathbf{u} = (77; 0; 36; 0; 0)^T \,\mathrm{PgC}\,\mathrm{yr}^{-1} \qquad (50)$$

and

$$B_\xi = \xi \begin{pmatrix} -77/37 & 0 & 0 & 0 & 0 \\ 31/37 & -31/452 & 0 & 0 & 0 \\ 0 & 0 & -36/69 & 0 & 0 \\ 21/37 & 15/452 & 12/69 & -48/81 & 0 \\ 0 & 2/452 & 6/69 & 3/81 & -11/1,121 \end{pmatrix} \text{yr}^{-1}, \quad (51)$$

where the numbers are chosen as in Thompson and Randerson (1999). The input vector is expressed in units of petagrams of carbon per year (PgC yr^{-1}) and the fractional transfer coefficients in units of per year (yr^{-1}). Because B_ξ is a lower triangular matrix, the model contains no feedbacks. For every value of ξ the system has a different steady state (Fig. 3, panel (a)). The higher the value of ξ, the faster the system, which makes the mean transit time (panel (b)) decrease, and because of shorter paths also the path entropy (panel (d)) decreases. Since ξ has no impact on the structure of the model, the mean number of jumps (panel (c)) remains unaffected. This can also be seen from the solid line marked by squares in panel (d)). It represents the part of the path entropy related to jump-associated uncertainties (Eq. (37)). The solid line marked by circles represents the part of the path entropy related to sojourn-associated uncertainties (Eq. (38)), which as a weighted average of one-pool entropies decreases similar to the entropy of an exponential distribution with increasing rate parameter λ (Fig. 1, panel (b)). The two parts together constitute the path entropy as represented by the unmarked solid line.

The entropy rate per unit time (panel (e)) increases until $\xi \approx 6$ and decreases afterwards, because with increasing system speed the uncertainty associated to sojourn times more and more dominates the uncertainty associated to jumps. While the uncertainty associated to jumps averaged over the path length increases because the total jump uncertainty is constant (see solid line marked with squares in panel (d) and the mean path length decreases (panel (b)), the sojourn-associated uncertainty decreases with increasing system speed for $\xi > 6$ similar to the entropy rate of a Poisson pro-

cess with intensity rate $\lambda > 1$. The entropy rate per jump (panel f) decreases with increasing ξ, because the path entropy of the system decreases.

Dashed lines in panels (d)–(f) show the respective entropy values for a one-pool system $M_\lambda = M((77 + 36)\,\mathrm{PgC\,yr}^{-1}, -\lambda)$ with the same mean transit time, that is $\lambda^{-1} = \mathbb{E}\left[\mathscr{T}_\xi\right]$. The solid and dashed lines intersect at $\xi \approx 4.31$ in panels (d) and (e). Before this break-even point the path of the Emanuel model is harder to predict than the path (that is the exit time of the particle) of a one-pool model with the same mean transit time. After this point of break even, the path of the Emanuel model with five compartments is easier to predict than only the transit time in a one-pool model. The reason is that as the system becomes faster, the differential entropy of the sojourn times in slow pools decreases so fast that at some point the sojourn times in slow pools visited by few particles becomes rather unimportant. The one-pool model's path becomes relatively harder to predict because it puts too much weight on a small amount of slowly cycling particles.

Note that there is no point in comparing jump-associated uncertainties (square-marked lines) with one-pool entropies (dashed lines), because the former are discrete entropies and the latter differential entropies. Comparison of a differential entropy with another quantity becomes only reasonable if a second differential entropy is involved as is true for the path entropy or the entropy rates θ and θ_J (unmarked solid lines). Hence, square- and circle-marked lines assist in understanding the composition of the entropies of the multi-pool system, and only the composition of the two can then be compared to the one-pool entropy rate.

3.3 A nonlinear autonomous soil organic matter decomposition model

Consider the nonlinear two-compartment model $M_\varepsilon = M(\mathbf{u}, \mathrm{B}_\varepsilon)$ described by Wang et al. (2014) which is used to represent the dynamics of microbes and carbon sub-
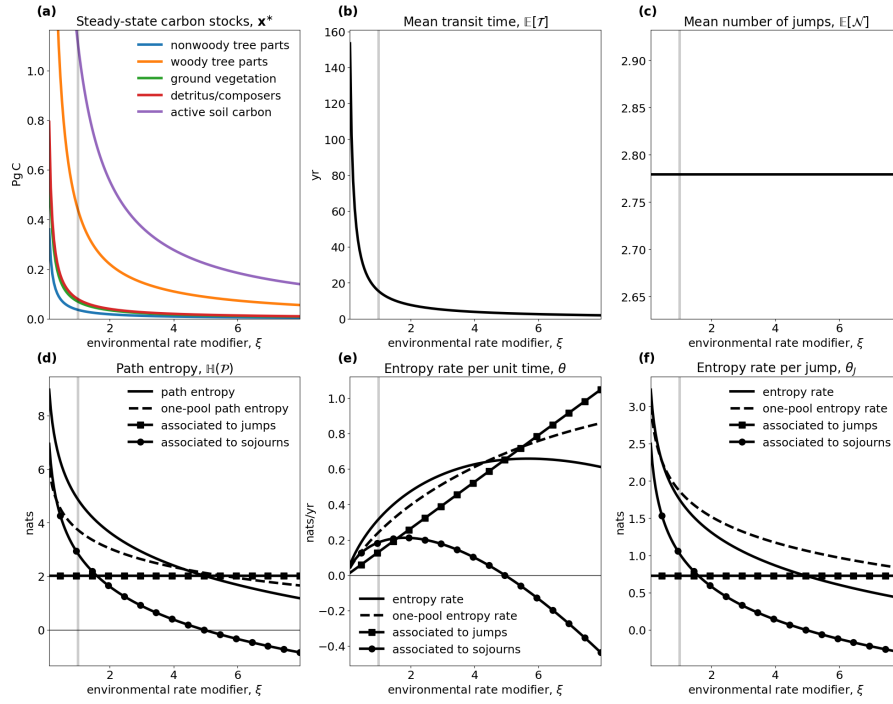
Fig. 3 (a) Equilibrium carbon stocks and (b)–(f) entropy related quantities of the global carbon cycle model introduced by Emanuel et al. (1981) in dependence on the environmental rate coefficient ξ

strates in soils (Fig. 4). Its ODE system is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} C_s \\ C_b \end{pmatrix}(t) = \begin{pmatrix} -\lambda(\mathbf{x}(t)) & \mu_b \\ \varepsilon\lambda(\mathbf{x}(t)) & -\mu_b \end{pmatrix} \begin{pmatrix} C_s \\ C_b \end{pmatrix} + \begin{pmatrix} F_{\mathrm{NPP}} \\ 0 \end{pmatrix}, \tag{52}$$

where $\mathbf{x}(t) = (C_s, C_b)^T(t)$. We denote by C_s and C_b soil organic carbon and soil microbial biomass carbon ($\mathrm{gC\,m^{-2}}$), respectively, by ε the carbon use efficiency or fraction of assimilated carbon that is converted into microbial biomass (unit-less), by μ_b the turnover rate of microbial biomass per year ($\mathrm{yr^{-1}}$), by F_{NPP} the carbon influx into the soil ($\mathrm{gC\,m^{-2}\,yr^{-1}}$), and by V_s and K_s the maximum rate of soil carbon assimilation per unit microbial biomass per year ($\mathrm{yr^{-1}}$) and the half-saturation constant for soil carbon assimilation by microbial biomass ($\mathrm{gC\,m^{-2}}$), respectively.
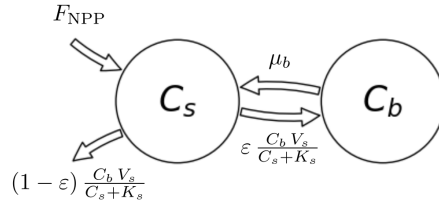
Fig. 4 Scheme of the nonlinear autonomous carbon cycle model introduced by Wang et al. (2014) with two compartments: substrate organic carbon (C_s) and microbial biomass (C_b)

We consider the model in equilibrium, that is $\mathbf{x}(t) = \mathbf{x}^* = (C_s^*, C_b^*)^T$ with

$$C_s^* = \frac{K_s}{\frac{V_s \varepsilon}{\mu_b} - 1} \quad \text{and} \quad C_b^* = \frac{F_{\mathrm{NPP}}}{\mu_b \left(-1 + \frac{1}{\varepsilon}\right)}. \tag{53}$$

The equilibrium stocks depend on the carbon use efficiency ε and so does the compartmental matrix $B = B_\varepsilon$, because

$$\lambda(\mathbf{x}) = \frac{C_b V_s}{C_s + K_s}. \tag{54}$$

From Wang et al. (2014) we take the parameter values $\mu_b = 4.38\,\mathrm{yr}^{-1}$, $F_{\mathrm{NPP}} = 345.00\,\mathrm{gC\,m}^{-2}\,\mathrm{yr}^{-1}$, and $K_s = 53{,}954.83\,\mathrm{gC\,m}^{-2}$. Since the description of V_s is missing in the original publication, we let it be equal to $59.13\,\mathrm{yr}^{-1}$ to approximately meet the given steady-state contents $C_s^* = 12{,}650.00\,\mathrm{gC\,m}^{-2}$ and $C_b^* = 50.36\,\mathrm{gC\,m}^{-2}$ for the original value $\varepsilon = 0.39$. Otherwise we leave the carbon use efficiency ε as a free parameter.

In contrast to the system from the first example, this system exhibits a feedback. This feedback results from dead soil microbial biomass being considered as new soil organic matter. The feedback can also be recognized by noting that B is not triangular. For every value of ε the system has a different steady state (Fig. 5, panel (a)). The higher the value of ε, the lower the equilibrium substrate organic carbon and the higher the microbial biomass carbon. Caused by the model's nonlinearity expressed in Eq. (54), the system speed increases and the mean transit time goes down (panel (b)) with increasing ε. At the same time, higher carbon use efficiency increases the probability of each carbon atom to be reused more often, hence the mean number of

jumps increases (panel (c)), making the entropy rate per jump decrease (panel (f)).
Even though the average paths become shorter, with increasing carbon use efficiency
the path entropy increases as well for most values of ε. This has two reasons. First,
the mean uncertainty of where to jump from C_s increases, this uncertainty decreases
then for $\varepsilon > 0.5$ (solid line marked by squares in panel (f)). Second, the rate $-B_{11}$ of
leaving the substrate pool is increasing and smaller than 1. The corresponding Poisson
process reaches its maximum entropy rate at an intensity rate equal to 1 (Fig. 1, panel
(c)), which corresponds to $\varepsilon \approx 0.926$. This is also reflected in the entropy rate per unit
time (panel (e)). The maximum does not exactly occur at $\varepsilon = 0.926$, because the times
that the particle stays in the different pools also depends on ε. For ε approaching 1,
both the path entropy and the entropy rate rapidly decline as the sojourn-associated
uncertainties (solid lines with circle markers) decline sharply because of a nonlinear
increase of the rate $-B_{11}$ of soil organic carbon turnover.

Considering a one-pool system $M_\lambda = M(345.00\,\mathrm{gC\,m^{-2}\,yr^{-1}}, -1/\mathbb{E}[\mathscr{T}_\varepsilon])$ with
the same mean transit time, we recognize only small sensitivity of the entropies on ε,
because the contrary effects on path length and jump- and sojourn-associated uncer-
tainties mostly balance out (dashed lines in panels d–f).

3.4 Model identification via Maxent

The following example is inspired by Anderson (1983, Example 16 C). It shows how
MaxEnt can help take a decision which model to use if not all parameters can be
uniquely determined from the transfer function $\widehat{\Psi}$. We are interested in determining
the entries of the compartmental matrix B belonging to the 2-dimensional equilibrium
compartmental system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) + \begin{pmatrix} 1 \\ 0 \end{pmatrix}\mathrm{gC\,yr^{-1}}, \quad t > 0. \tag{55}$$
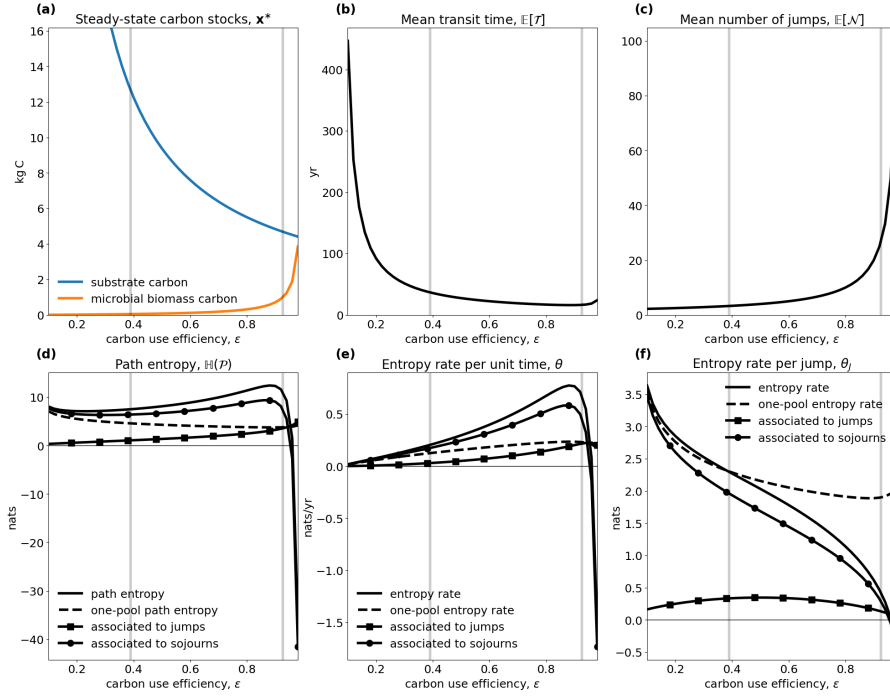
Fig. 5 (a) Equilibrium carbon stocks and (b)–(f) entropy related quantities of the global carbon cycle model introduced by Wang et al. (2014) in dependence on the microbial carbon use efficiency ε

We immediately notice $\mathbf{u} = (1,0)^T \, \mathrm{gC} \, \mathrm{yr}^{-1}$ and $A = I$. Further, we decide to measure the contents of compartment 1 such that $C = (1,0)$. We recall $z_j = -\sum_{i=1}^d B_{ij}$ and obtain $z_1 = -B_{11} - B_{21}$ and $z_2 = -B_{22} - B_{12}$. The real-valued transfer function is then given by

$$\widehat{\Psi}(s) = \frac{s + \gamma_1}{s^2 + \gamma_2 \, s + \gamma_3}, \tag{56}$$

where

$$
\begin{aligned}
\gamma_1 &= B_{12} + z_2, \\
\gamma_2 &= B_{21} + z_1 + B_{12} + z_2, \\
\gamma_3 &= z_1 \, B_{12} + z_1 \, z_2 + B_{21} \, z_2.
\end{aligned}
\tag{57}
$$

We assume that $\widehat{\Psi}$ is known from measurements, that is, γ_1, γ_2, and γ_3 are known impulse response parameters. We have the four unknown parameters B_{11}, B_{12}, B_{21},

520   and B_{22}, or equivalently, B_{12}, B_{21}, z_1, and z_2, but only three equations to determine

521   them. Consequently, the system is nonidentifiable and there remains a class $\mathcal{M}$ of

522   models which all satisfy Eq. (57). Which model out of $\mathcal{M}$ are we going to select

523   now?

524   Here, MaxEnt comes into play. We intend to select the model that best represents

525   the information given by our measurement data. We have to find $M^* = M(\mathbf{u}, B^*)$ such

526   that

$$M^* = \underset{M \in \mathcal{M}}{\arg\max} \, \theta(M). \tag{58}$$

527   We maximize the entropy rate per unit time here instead of the path entropy, because

528   by slowing down the model we could potentially increase its mean transit time and

529   with it its path entropy indefinitely.

530   Let us turn to a numerical example in which we suppose to be given $\gamma_1 = 3 \, \text{yr}^{-1}$,

531   $\gamma_2 = 5 \, \text{yr}^{-1}$, and $\gamma_3 = 4 \, \text{yr}^{-1}$. A nonlinear optimization algorithm with the arbitrarily

532   chosen initial values $B_{12} = 3 \, \text{yr}^{-1}$, $B_{21} = 0 \, \text{yr}^{-1}$, $z_1 = 1 \, \text{yr}^{-1}$, and $z_2 = 1 \, \text{yr}^{-1}$ ends

533   approximately with the terminal compartmental matrix

$$B^* \approx \begin{pmatrix} -2.00 & 1.90 \\ 1.05 & -3.00 \end{pmatrix} \text{yr}^{-1} \tag{59}$$

534   and the terminal entropy rate per unit time $\theta(M^*) \approx 1.92 \, \text{nats} \, \text{yr}^{-1}$. Unfortunately,

535   this local maximum solution it is not guaranteed to be a global maximum entropy

536   model in $\mathcal{M}$.

537   The nonidentifiability of the model from $\widehat{\Psi}$ alone is underlined by the fact that

538   another system $\widetilde{M} = M(\mathbf{u}, \widetilde{B}) \in \mathcal{M}$ with

$$\widetilde{B} = \begin{pmatrix} -2.00 & 2.00 \\ 1.00 & -3.00 \end{pmatrix} \text{yr}^{-1} \tag{60}$$

results in the same transfer function, but a different entropy rate per unit time, that is, $\theta(\widetilde{M}) \approx 1.90\,\mathrm{nats}\,\mathrm{yr}^{-1}$. Following MaxEnt, in this example we select M^*, which is found by a nonlinear optimization process whose evolution is depicted by the solid line in Fig. 6. Peaks higher than the terminal value show attempts of the optimization algorithm that do not perfectly satisfy all constraints and are rejected. The dashed line shows the entropy rate per unit time of model $\widetilde{M}$.
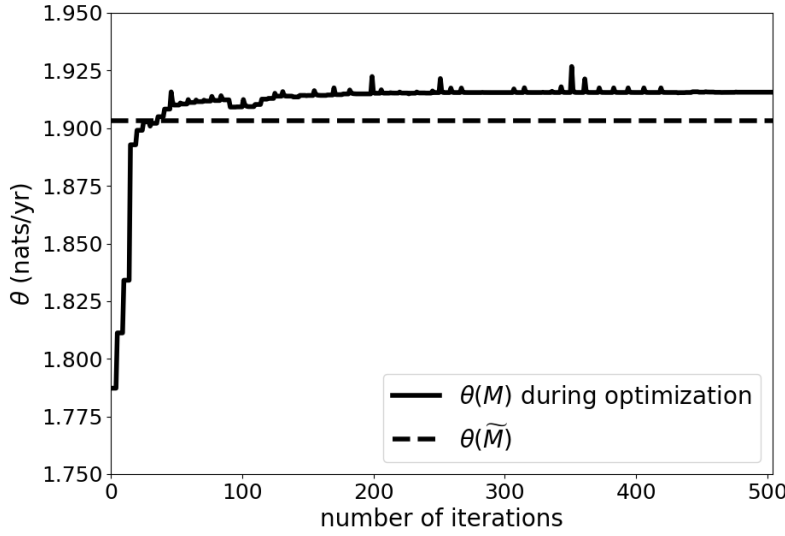
Fig. 6 Evolution of the entropy rate per unit time of system (55) during the nonlinear optimization process for model identification

4 Discussion

Based on the path that a particle takes through a compartmental system, we introduced three types of entropy based on Shannon's information theory. The entropy of the particle's entire path through the system is the central concept, and the entropy rates per unit time and per jump are consistently derived from it. Even though we call $\mathbb{H}(\mathscr{P})$ path entropy and identify models by maximizing it, it is different from the con-

cept of path entropy as treated in the context of maximum caliber (MaxCal) (Jaynes 1985). We maximize here the Shannon entropy of a single particle's microscopic path through a compartmental system by means of an absorbing continuous-time Markov chain whose transition probabilities are already determined by the macroscopic equilibrium state of the system. As discussed by Pressé et al. (2013), MaxCal interprets the path entropy as a macroscopic system property to be maximized in order to identify a time-dependent trajectory of the entire dynamical system, not just one single particle.

In the field of soil carbon cycle modeling, Ågren (2021) recently applied the maximum entropy principle to identify the distribution of soil carbon qualities within the framework of the continuous-quality theory. Given only the nonnegative mean quality, an application of MaxEnt leads to an exponential quality distribution, because under these circumstances the exponential distribution is the maximum entropy distribution. The path entropy generalizes this approach to several interconnected compartments and jumps between them, while each sojourn time in a compartment is exponentially distributed.

From the simple examples in Sect. 3.1 we can see that models can be ordered differently in terms of uncertainty, depending on whether the interest is in the uncertainty of the entire path or in some average uncertainty rate. For applications of MaxEnt without restrictions on the transit time, it is often useful to maximize an entropy rate because by slowing the system down more and more, the path entropy can potentially be increased indefinitely and there is no way to find a maximum path entropy model.

Usually, entropy is maximized when the system is highly symmetric. This is indicated by the Bernoulli entropy (Fig. 1, panel (a)) and supported by Example 1. Intuitively, this result is obvious. If a system has high symmetry, a particle is equally likely to jump among different pools, and the Poisson process with intensity rate 1 is the one with maximum entropy rate. Furthermore, the resulting rates $z_j = 1/\mathbb{E}[\mathscr{T}]$

of leaving the system are chosen such that the mean transit time constraint is fulfilled. In Example 2, the symmetry is broken by the additional restriction of a given steady-state vector. Consequently, $\mathbb{H}(M_2) \leq \mathbb{H}(M_1)$.

When we compute entropy values for actual carbon-cycle models (Sects. 3.2 and 3.3), we note that environmental or biochemical factors impact the model entropies. Furthermore, we see that for sufficiently fast systems, a multi-pool model has lower entropy than a one-pool model with the same system speed. The one-pool system might put too much weight on the uncertainties of a small amount of slow-cycling particles, while the more detailed multi-pool model focuses more on the small uncertainties of the major amount of fast-cycling particles. The path of a detailed model that separates fast from slow paths is then even easier to predict than a one-pool model path, even though the detailed model's path looks more complicated. However, detailed paths of slow-cycling systems are harder to predict than just the exit-time in a one-pool equivalent.

The example of model identification by MaxEnt in Sect. 3.4 shows a major difference to the more artificial previous maximum entropy examples. The given constraints do not tell us enough about the structure of the model class $\mathscr{M}$ to ensure that an identified local maximum is also a global maximum. It might be possible that with different initial values the optimization algorithm finds another local maximum model with higher entropy rate. This example is only supposed to give a first impression of how the maximum entropy principle can be used in combination with entropy rates or path entropy in similar situations. Practical examples usually have a high level of complexity such that existence and uniqueness of a maximum entropy model have to be studied on a case-by-case basis.

5 Conclusions

Information content and complexity of autonomous compartmental systems in equilibrium can be assessed by the entropy of the path of particles traveling through a system of interconnected compartments. When a particle moves through a compartmental system, it creates a path from the time of its entry until the time of its exit. This path can be described in three ways: (1) as a random variable in the path space; (2) as a continuous-time stochastic process representing the visited compartments; (3) as a discrete sequence of pairs consisting of visited compartments and associated sojourn times. Based on these three ways, we introduced for systems in equilibrium (1) the entropy of the entire path, (2) the entropy rate per unit time, and (3) the entropy rate per jump. These three different entropies allow to quantify how difficult it is to predict the path of particles entering a compartmental system, serving as a measure of complexity and information content. With these measures, it is thus possible to apply maximum entropy principles to compartmental systems in equilibrium in order to address problems of equifinality in model selection.

Although the path entropy concept developed here only applies to systems in equilibrium, it sets the foundation for future research on systems out of equilibrium. This could be done by building on the concept of the entropy rate per unit time as an instantaneous uncertainty and interpreting non-autonomous compartmental systems as inhomogeneous Markov chains. This would allow an extension of MaxCal applied only to the inhomogeneous embedded jump chain as done by Ge et al. (2012).

By introducing the concept of path entropy to compartmental systems, we made a first crucial step toward a quantification of information content in models that can be compared to other methods to obtain information content from observations. Using entropy measures in both models and observations, we could potentially advance toward better methods for model selection applying the maximum entropy principle.

footnotesize

References

Ågren G I (2021) Investigating soil carbon diversity by combining the maximum entropy principle with the q model. Biogeochemistry 153(1):85–94

Albert A (1962) Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process. The Annals of Mathematical Statistics 33(2):727–753

Anderson D H (1983) Compartmental modeling and tracer kinetics, volume 50. Springer Science & Business Media

Bad Dumitrescu M E (1988) Some informational properties of markov pure-jump processes. Časopis pro pěstování matematiky 113(4):429–434

Bellman R, Åström K J (1970) On structural identifiability. Mathematical Biosciences 7(3-4):329–339

Bolin B, Rodhe H (1973) A note on the concepts of age distribution and transit time in natural reservoirs. Tellus 25(1):58–62

Bonchev D, Buck G A (2005) Quantitative Measures of Network Csomplexity. In Complexity in Chemistry, Biology, and Ecology, Springer, 191–235

Cover T M, Thomas J A (2006) Elements of Information Theory. Wiley, 2 edition

Dehmer M, Mowshowitz A (2011) A history of graph entropy measures. Information Sciences 181(1):57–78

Doob J L (1953) Stochastic Processes, volume 7. Wiley, New York

Emanuel W R, Killough G G, Olson J S (1981) Modelling the Circulation of Carbon in the World's Terrestrial Ecosystems. In Carbon Cycle Modelling, SCOPE 16, John Wiley and Sons, 335–353

655 Eriksson E (1971) Compartment Models and Reservoir Theory. Annual Review of
656   Ecology and Systematics 2:67–84

657 Gaspard P, Wang X J (1993) Noise, chaos, and (ε, τ)-entropy per unit time. Physics
658   Reports 235(6):291–343

659 Ge H, Pressé S, Ghosh K, Dill K A (2012) Markov processes follow from the princi-
660   ple of maximum caliber. The Journal of Chemical Physics 136(6):064108

661 Girardin V (2004) Entropy Maximization for Markov and Semi-Markov Processes.
662   Methodology and Computing in Applied Probability 6(1):109–127

663 Girardin V, Limnios N (2003) On the Entropy for Semi-Markov Processes. Journal
664   of Applied Probability 40(4):1060–1068

665 Haddad W M, Chellaboina V, Hui Q (2010) Nonnegative and Compartmental Dy-
666   namical Systems. Princeton University Press

667 Jacquez J A, Simon C P (1993) Qualitative theory of compartmental systems. Siam
668   Review 35(1):43–79

669 Jaynes E T (1957a) Information Theory and Statistical Mechanics. Physical Review
670   106(4):620–630

671 Jaynes E T (1957b) Information Theory and Statistical Mechanics. ii. Physical Re-
672   view 108(2):171–190

673 Jaynes E T (1985) Macroscopic prediction. In Complex Systems – Operational Ap-
674   proaches in Neurobiology, Physics, and Computers, Springer, 254–269

675 Metzler H, Müller M, Sierra C A (2018) Transit-time and age distributions for nonlin-
676   ear time-dependent compartmental systems. Proceedings of the National Academy
677   of Sciences 115(6):1150–1155

678 Metzler H, Sierra C A (2018) Linear autonomous compartmental models as
679   continuous-time Markov chains: Transit-time and age distributions. Mathematical
680   Geosciences 50(1):1–34

681 Neuts M F (1981) Matrix-geometric solutions in stochastic models: An algorithmic
682   approach. The Johns Hopkins University Press

Norris J R (1997) Markov Chains. Cambridge University Press

Pesin Y B (1977) Characteristic Lyapunov exponents and smooth ergodic theory. Uspekhi Matematicheskikh Nauk 32(4):55–112

Pressé S, Ghosh K, Lee J, Dill K A (2013) Principles of maximum entropy and maximum caliber in statistical physics. Reviews of Modern Physics 85(3):1115

Rasmussen M, Hastings A, Smith M J, Agusto F B, Chen-Charpentier B M, Hoffman F M, Jiang J, Todd-Brown K E O, Wang Y, Wang Y P, Luo Y (2016) Transit times and mean ages for nonautonomous and autonomous compartmental systems. Journal of Mathematical Biology 73(6-7):1379–1398

Shannon C E, Weaver W (1949) The Mathematical Theory of Communication. The University of Illinois Press, Urbana

Sierra C A, Müller M, Metzler H, Manzoni S, Trumbore S E (2016) The muddle of ages, turnover, transit, and residence times in the carbon cycle. Global Change Biology in print

Thompson M V, Randerson J T (1999) Impulse response functions of terrestrial carbon cycle models: method and application. Global Change Biology 5(4):371–394

Trucco E (1956) A note on the information content of graphs. Bulletin of Mathematical Biology 18(2):129–135

Walter G G (1986) Size identifiability of compartmental models. Mathematical Biosciences 81(2):165–176

Walter G G, Contreras M (1999) Compartmental Modeling with Networks. Birkhäuser

Wang Y P, Chen B C, Wieder W R, Leite M, Medlyn B E, Rasmussen M, Smith M J, Agusto F B, Hoffman F, Luo Y Q (2014) Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. Biogeosciences 11(7):1817–1831

A Basic ideas of Shannon information entropy

We introduce basic concepts of information entropy along the lines of Cover and Thomas (2006). There are two concepts of entropy of a random variable, depending on whether the random variable has a discrete or a continuous distribution.

Definition A.1 (1) Let Y_d be a discrete real-valued random variable with range R_d and probability mass function p. The "Shannon information entropy" or "Shannon entropy" or "information entropy", or simply "entropy" of Y_d is defined by

$$\mathbb{H}(Y_d) = -\sum_{y \in R_d} p(y) \log p(y) = -\mathbb{E}\left[\log p(Y_d)\right]. \tag{A.1}$$

By convention, $0 \log 0 := 0$.

(2) Let Y_c be a continuous real-valued random variable with range R_c and probability density function f. Then the "differential entropy" or simply "entropy" of Y_c is defined by

$$\mathbb{H}(Y_c) = -\int_{R_c} f(y) \log f(y) \, \mathrm{d}y = -\mathbb{E}\left[\log f(Y_c)\right]. \tag{A.2}$$

Remark A.1 Depending on the base of the logarithm, the unit of the entropy changes. For base 2, the unit is called "bits" and for the natural logarithm with base e, the unit is called "nats". If not stated differently, we use the value e as logarithmic base, that is, we use the natural logarithm.

The entropy $\mathbb{H}(Y)$ of a random variable Y has two intertwined interpretations. On the one hand, $\mathbb{H}(Y)$ is a measure of uncertainty, that is, a measure of how difficult it is to predict the outcome of a realization of Y. On the other hand, $\mathbb{H}(Y)$ is also a measure of the information content of Y, that is, a measure of how much information we gain once we learn about the outcome of a realization of Y. It is important to note that, even though their definitions and information theoretical interpretations are quite similar, the Shannon- and the differential entropy are of different nature. The Shannon entropy is always nonnegative, whereas the differential entropy can have negative values. Consequently, the Shannon entropy is an absolute measure of information and makes sense in its own right. The differential entropy, however, is not an absolute information measure. Hence, the differential entropy of a random variable makes sense only in comparison with the differential entropy of another random variable.

Panel (a) of Fig. 1 depicts the Shannon entropy of a Bernoulli random variable Y_d with $\mathbb{P}(Y_d = 1) = 1 - \mathbb{P}(Y_d = 0) = p \in [0,1]$. This random variable could represent the outcome of a coin toss. We can see that the entropy is low when p is close to 0 or 1. In these cases, we have some information that the coin is biased, and hence we have a preference if we guess the outcome. The entropy is maximum if the coin is

fair ($p = 1/2$), since we have no additional information about the outcome of the coin toss. The Shannon entropy of Y_d is

$$\mathbb{H}(Y_d) = -p \log p - (1-p) \log(1-p). \tag{A.3}$$

Panel (b) of Fig. 1 shows the differential entropy of an exponentially distributed random variable $Y_c \sim \mathrm{Exp}(\lambda)$ with rate parameter $\lambda > 0$, probability density function $f(y) = \lambda\, e^{-\lambda y}$ for $y \geq 0$, and $\mathbb{E}[Y_c] = \lambda^{-1}$. We can imagine it to represent the duration of stay of a particle in a well-mixed compartment in a linear autonomous compartmental system, where λ is the total outflow rate from the compartment. The higher the outflow rate, the likelier an early exit of the particle, and the easier it is to predict the moment of exit. Hence, the differential entropy

$$\mathbb{H}(Y_c) = 1 - \log \lambda \tag{A.4}$$

decreases with increasing λ.

Definition A.2 Let Y_1, Y_2 be two discrete random variables with joint probability mass function p and ranges R_1 and R_2, respectively. The "joint entropy" of Y_1 and Y_2 is defined by

$$\mathbb{H}(Y_1, Y_2) = -\sum_{y_1 \in R_1} \sum_{y_2 \in R_2} p(y_1, y_2) \log p(y_1, y_2) = -\mathbb{E}[\log p(Y_1, Y_2)]. \tag{A.5}$$

Note that the joint entropy is symmetric, that is, $\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_2, Y_1)$. Furthermore, $\mathbb{H}(Y_1, Y_2) \leq \mathbb{H}(Y_1) + \mathbb{H}(Y_2)$ with equality if Y_1 and Y_2 are independent.

Definition A.3 Let Y_1 and Y_2 be two discrete random variables with joint probability mass function p. Furthermore, let p_2 denote the probability mass function of Y_2 and denote by $p(y_1 \,|\, y_2)$ the conditional probability $\mathbb{P}(Y_1 = y_1 \,|\, Y_2 = y_2)$.

Then the "conditional entropy" of Y_1 given Y_2 is defined by

$$
\begin{aligned}
\mathbb{H}(Y_1 \,|\, Y_2) &= \sum_{y_2 \in R_2} \mathbb{H}(Y_1 \,|\, Y_2 = y_2)\, p_2(y_2) \\
&= -\sum_{y_2 \in R_2} p_2(y_2) \sum_{y_1 \in R_1} p(y_1 \,|\, y_2) \log p(y_1 \,|\, y_2) \\
&= -\sum_{y_2 \in R_2} \sum_{y_1 \in R_1} p(y_1, y_2) \log p(y_1 \,|\, y_2) \\
&= -\mathbb{E}[\log p(Y_1 \,|\, Y_2)].
\end{aligned}
\tag{A.6}
$$

Note that $\mathbb{H}(Y_1 \,|\, Y_2) \leq \mathbb{H}(Y_1)$ with equality if Y_1 and Y_2 are independent. The joint entropy of two random variables is the entropy of one variable plus the conditional entropy of the other. This is expressed in

$$\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_2) + \mathbb{H}(Y_1 \,|\, Y_2). \tag{A.7}$$

Let Y_3 be a third discrete random variable. Then

$$\mathbb{H}(Y_1, Y_2 \,|\, Y_3) = \mathbb{H}(Y_1 \,|\, Y_3) + \mathbb{H}(Y_2 \,|\, Y_1, Y_3). \tag{A.8}$$

Let $Y_1, Y_2, \ldots, Y_n$ be discrete random variables. By repeated application of Eq. (A.7) and Eq. (A.8), we obtain the "chain rule"

$$\mathbb{H}(Y_1, Y_2, \ldots, Y_n) = \sum_{k=1}^{n} \mathbb{H}(Y_k \,|\, Y_{k-1}, \ldots, Y_1). \tag{A.9}$$

Remark A.2 We defined the joint- and conditional entropy for discrete random variables only. Analogous definitions can be made for continuous random variables. Also the chain rule holds for differential entropy.

Definition A.4 The "entropy rate" of a discrete-time stochastic process $Y = (Y_n)_{n \in \mathbb{N}}$ is defined by

$$\theta(Y) = \lim_{n \to \infty} \frac{1}{n} \mathbb{H}(Y_1, Y_2, \ldots, Y_n) = -\frac{1}{n} \mathbb{E}\left[\log p_n(Y_1, Y_2, \ldots, Y_n)\right] \tag{A.10}$$

if the limit exists. Here, p_n denotes the joint probability mass function of $Y_1, Y_2, \ldots, Y_n$.

The discrete-time entropy rate describes the long-term average increase of the processes' entropy per time step. The statements of the following lemma are proven in Cover and Thomas (2006, Theorem 4.2.1).

Lemma A.1 For a stationary discrete-time stochastic process $Y = (Y_n)_{n \in \mathbb{N}}$, the entropy rate is

$$\theta(Y) = \lim_{n \to \infty} \mathbb{H}(Y_n \,|\, Y_{n-1}, \ldots, Y_1). \tag{A.11}$$

Consequently, if Y is a stationary discrete-time Markov chain, its entropy rate is

$$\theta(Y) = \mathbb{H}(Y_2 \,|\, Y_1). \tag{A.12}$$

According to Bad Dumitrescu (1988) and Girardin and Limnios (2003), we can also define the entropy rate for continuous-time processes. To that end, we first define the entropy on a finite time interval.

Definition A.5 The "finite-time entropy" of the continuous-time stochastic process $Z = (Z_t)_{t \geq 0}$ until $T \geq 0$ is defined as

$$\mathbb{H}_T(Z) = -\int f_T(z) \log f_T(z) \, d\mu_T(z), \tag{A.13}$$

where f_T is the probability density function of $(Z_t)_{0 \leq t \leq T}$ with respect to some reference measure μ_T, if it exists.

Definition A.6 The "entropy rate" of a continuous-time stochastic process $Z = (Z_t)_{t \geq 0}$ is defined by

$$\theta(Z) = \lim_{T \to \infty} \frac{1}{T} \mathbb{H}_T(Z) \tag{A.14}$$

if the limit exists.

B Proves of the MaxEnt examples

Recall that the path entropy of a linear autonomous compartmental system $M = M(\mathbf{u}, \mathrm{B})$ is given by

$$
\begin{aligned}
\mathbb{H}(M) &= \mathbb{H}(X) \\
&= -\sum_{i=1}^{d} \beta_i \log \beta_i + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^{d} B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j) \right].
\end{aligned}
\tag{B.1}
$$

In order to obtain maximum entropy models under simple constraints, we now adapt ideas of Girardin (2004).

Proposition B.1 Consider the set $\mathscr{M}_1$ of compartmental systems in equilibrium given by Eq. (8) with a predefined nonzero input vector $\mathbf{u}$, a predefined mean transit time $\mathbb{E}[\mathscr{T}]$, and an unknown steady-state vector comprising nonzero components. The compartmental system $M_1^* = M(\mathbf{u}, \mathrm{B}^*)$ with

$$
\mathrm{B}^* = \begin{pmatrix}
-\lambda & 1 & \cdots & 1 \\
1 & -\lambda & 1 \cdots & 1 \\
\vdots & & \ddots & \vdots \\
1 & \cdots & 1 & -\lambda
\end{pmatrix},
\tag{B.2}
$$

where $\lambda = d - 1 + 1/\mathbb{E}[\mathscr{T}]$, is the maximum entropy model in $\mathscr{M}_1$.

Proof We can express the constraint $\mathbb{E}[\mathscr{T}] = \|\mathbf{x}^*\|/\|\mathbf{u}\|$ by

$$
C_1 = \frac{1}{\|\mathbf{u}\|} \sum_{j=1}^{d} x_j^* - \mathbb{E}[\mathscr{T}] = 0.
\tag{B.3}
$$

From the steady-state formula $\mathbf{x}^* = -\mathrm{B}^{-1} \mathbf{u}$, we obtain another set of d constraints, which we can describe by

$$
\frac{1}{\|\mathbf{u}\|} (\mathrm{B}\mathbf{x}^*)_i = -\beta_i, \quad i = 1, 2, \ldots, d.
\tag{B.4}
$$

We rewrite the left hand side as

$$
\begin{aligned}
\frac{1}{\|\mathbf{u}\|} (\mathrm{B}\mathbf{x}^*)_i &= \frac{1}{\|\mathbf{u}\|} \sum_{j=1}^{d} B_{ij} x_j^* = \frac{1}{\|\mathbf{u}\|} \left(\sum_{j=1, j \neq i}^{d} B_{ij} x_j^* + B_{ii} x_i^* \right) \\
&= \frac{1}{\|\mathbf{u}\|} \sum_{j=1, j \neq i}^{d} B_{ij} x_j^* - \frac{1}{\|\mathbf{u}\|} x_i^* \left(\sum_{k=1, k \neq i}^{d} B_{ki} + z_i \right),
\end{aligned}
\tag{B.5}
$$

785 which leads to the constraints

$$C_{2,i} = \frac{1}{\|\mathbf{u}\|} \sum_{j=1, j \neq i}^{d} B_{ij} x_j^* - \frac{1}{\|\mathbf{u}\|} x_i^* \left(\sum_{k=1, k \neq i}^{d} B_{ki} + z_i \right) + \beta_i = 0, \quad i \in S. \tag{B.6}$$

786 The Lagrangian is now given by

$$L = \mathbb{H}(X) + \gamma_0 C_1 + \sum_{i=1}^{d} \gamma_i C_{2,i} \tag{B.7}$$

787 and its partial derivatives with respect to B_{ij} $(i \neq j)$, z_j, and x_j^* by

$$\|\mathbf{u}\| \frac{\partial}{\partial B_{ij}} L = -x_j^* \log B_{ij} + \gamma_i x_j^* - \gamma_j x_j^*,$$
$$\|\mathbf{u}\| \frac{\partial}{\partial z_j} L = -x_j^* \log z_j - \gamma_j x_j^*, \tag{B.8}$$

788 and

$$\|\mathbf{u}\| \frac{\partial}{\partial x_j^*} L = \sum_{i=1, i \neq j}^{d} B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j)$$
$$+ \gamma_0 + \sum_{i=1, i \neq j}^{d} \gamma_i B_{ij} - \gamma_j \left(\sum_{k=1, k \neq j}^{d} B_{kj} + z_j \right), \tag{B.9}$$

789 respectively. Setting $\frac{\partial}{\partial B_{ij}} L = 0$ gives $B_{ij} = e^{\gamma_i - \gamma_j}$, and setting $\frac{\partial}{\partial z_j} L = 0$ gives $z_j = e^{-\gamma_j}$. We plug this into

790 $\frac{\partial}{\partial x_j^*} L = 0$ and get

$$0 = \sum_{i=1, i \neq j}^{d} e^{\gamma_i - \gamma_j} [1 - (\gamma_i - \gamma_j)] + e^{-\gamma_j} [1 - (-\gamma_j)]$$
$$+ \gamma_0 + \sum_{i=1, i \neq j}^{d} \gamma_i e^{\gamma_i - \gamma_j} - \gamma_j \left(\sum_{k=1, k \neq j}^{d} e^{\gamma_k - \gamma_j} + e^{-\gamma_j} \right) \tag{B.10}$$
$$= \sum_{i \neq j, i \neq j} e^{\gamma_i - \gamma_j} + e^{-\gamma_j} + \gamma_0.$$

791 Subtracting $e^{-\gamma_j}$ from both sides and multiplying with e^{γ_j} leads to

$$\gamma_0 e^{\gamma_j} + \sum_{i=1, i \neq j}^{d} e^{\gamma_i} = -1, \quad j = 1, 2, \ldots, d. \tag{B.11}$$

792 This is equivalent to the linear system $Y \mathbf{v} = -\mathbf{1}$ with

$$Y = \begin{pmatrix} \gamma_0 & 1 & \cdots & 1 \\ 1 & \gamma_0 & 1 \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & \gamma_0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} e^{\gamma_1} \\ e^{\gamma_2} \\ \vdots \\ e^{\gamma_d} \end{pmatrix}, \quad -\mathbf{1} = \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}. \tag{B.12}$$

The case $\gamma_0 = 1$ has no solution $\mathbf{v}$ since $e^{\gamma_i} > 0 > -1$. For $\gamma_0 \neq 1$ the matrix Y has a nonzero determinant which makes the system uniquely solvable. For symmetry reasons, $\gamma_i = \gamma_j =: \gamma$ for all $i,j = 1,2,\ldots,d$. Consequently, for $i \neq j$, we get $B_{ij} = 1$, and by summing Eq. (B.6) over $i \in S$,

$$
\begin{aligned}
0 = \|\mathbf{u}\| \sum_{i=1}^{d} C_{2,i} &= \sum_{i=1}^{d} \sum_{j=1,j\neq i}^{d} B_{ij} x_j^* - \sum_{i=1}^{d} x_i^* \left(\sum_{k=1,k\neq i}^{d} B_{ki} + z_i \right) - \|\mathbf{u}\| \\
&= -\sum_{i=1}^{d} x_i^* z_i - \|\mathbf{u}\|,
\end{aligned}
\tag{B.13}
$$

which can also be expressed by $\mathbf{z}^T \mathbf{x}^* = \|\mathbf{u}\|$. We simply plug in $z_i = e^{-\gamma}$ and get $e^{-\gamma}\|\mathbf{x}^*\| = \|\mathbf{u}\|$, which means $z_i = 1/\mathbb{E}[\mathscr{T}]$. Consequently,

$$
\mathbf{B}^* = \begin{pmatrix}
-\lambda & 1 & \cdots & 1 \\
1 & -\lambda & 1 \cdots & 1 \\
\vdots & & \ddots & \vdots \\
1 & \cdots & 1 & -\lambda
\end{pmatrix}
\tag{B.14}
$$

for $\lambda = d - 1 + 1/\mathbb{E}[\mathscr{T}]$. Since uniqueness of this solution follows from its construction, we remain with showing maximality. To this end, we split the entropy into to three parts, that is, $\mathbb{H}(X) = H_1 + H_2 + H_3$ with

$$
\begin{aligned}
H_1 &= -\sum_{i=1}^{d} \beta_i \log \beta_i, \\
H_2 &= \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} z_j (1 - \log z_j), \text{ and} \\
H_3 &= \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \sum_{i=1,i\neq j}^{d} B_{ij} (1 - \log B_{ij}).
\end{aligned}
\tag{B.15}
$$

The term H_1 is independent of B_{ij} and z_j for all $i,j \in S$ and $i \neq j$, and can thus be ignored. We denote by E the pool from which the particle exits from the system. Then we can use (Metzler and Sierra 2018, Sect. 5.3)

$$
\mathbb{P}(E = j) = \frac{z_j x_j^*}{\|\mathbf{u}\|}
\tag{B.16}
$$

to rewrite the second term as

$$
H_2 = \sum_{j=1}^{d} \mathbb{P}(E = j)(1 - \log z_j) = \sum_{j=1}^{d} \mathbb{H}(T_E \,|\, E = j)\,\mathbb{P}(E = j) = \mathbb{H}(T_E \,|\, E),
\tag{B.17}
$$

where T_E denotes the exponentially distributed sojourn time in E right before absorption. We see that H_2 becomes maximal if the knowledge of E contains no information about T_E. Hence, $z_j = z_i$ for $i,j \in S$. Since we need to ensure the systems' constraint on $\mathbb{E}[\mathscr{T}]$, we get $z_j = 1/\mathbb{E}[\mathscr{T}]$ for all $j \in S$.

808   In order to see that $B_{ij} = 1$ $(i \neq j)$ leads to maximal entropy, we first note that

$$H_3 = \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \sum_{i=1,i\neq j}^{d} 1 \cdot (1 - \log 1) = (d-1) \sum_{j=1}^{d} \mathbb{E}\left[O_j\right] = (d-1)\mathbb{E}[\mathscr{T}] \tag{B.18}$$

809   by Eq. (35). We now disturb B_{kl} for fixed $k,l \in S$ with $k \neq l$ by a sufficiently tiny ε, which may be
810   positive or negative. We define $B_{kl}(\varepsilon) := B_{kl} + \varepsilon$, and a change from λ_j to $\lambda_j(\varepsilon) := \lambda_j + \varepsilon > 0$ ensures
811    $z_j(\varepsilon) = z_j$, implying that the system's mean transit time remains unchanged, that is, $\mathbb{E}[\mathscr{T}(\varepsilon)] = \mathbb{E}[\mathscr{T}]$.
812   The ε-disturbed H_3 is given by

$$\begin{aligned}
H_3(\varepsilon) &= \sum_{j=1}^{d} \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1,i\neq j}^{d} 1 \cdot (1 - \log 1)\left(1 - \mathbb{1}_{\{i=k,j=l\}}\right) \\
&\quad + \frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|}(1+\varepsilon)[1 - \log(1+\varepsilon)] \\
&= \sum_{j=1}^{d} \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1,i\neq j}^{d}\left(1 - \mathbb{1}_{\{i=k,j=l\}}\right) + \frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|}(1-\delta)
\end{aligned} \tag{B.19}$$

813   for some $\delta > 0$ since the map $x \mapsto x(1 - \log x)$ has its global maximum at $x = 1$. Consequently,

$$\begin{aligned}
H_3(\varepsilon) \quad &= \left[\sum_{j=1}^{d} \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1,i\neq j}^{d} 1\right] - \delta\frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} \quad &= (d-1)\sum_{j=1}^{d} \mathbb{E}\left[O_j(\varepsilon)\right] - \delta\frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} \\
&= (d-1)\mathbb{E}[\mathscr{T}(\varepsilon)] - \delta\frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} \quad &= (d-1)\mathbb{E}[\mathscr{T}] - \delta\frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} \\
&< H_3.
\end{aligned} \tag{B.20}$$

814   Hence, disturbing B_{ij} away from 1 reduces the entropy of the system, and the proof is complete.

815   **Proposition B.2** Consider the set $\mathscr{M}_2$ of compartmental systems in equilibrium given by Eq. (8) with a
816   predefined nonzero input vector $\mathbf{u}$ and a predefined positive steady-state vector $\mathbf{x}^*$. The compartmental
817   system $M_2^* = M(\mathbf{u}, B^*)$ with $B^* = (B_{ij})_{i,j\in S}$ given by

$$B_{ij} = \begin{cases} \sqrt{\dfrac{x_i^*}{x_j^*}}, & i \neq j, \\ -\sum\limits_{k=1,k\neq j}^{d} \sqrt{\dfrac{x_k^*}{x_j^*}} - \dfrac{1}{\sqrt{x_j^*}}, & i = j, \end{cases} \tag{B.21}$$

818   is the maximum entropy model in $\mathscr{M}_2$.

819   *Proof* The mean transit time $\mathbb{E}[\mathscr{T}] = \|\mathbf{x}^*\|/\|\mathbf{u}\|$ of the system is fixed. Hence, the Lagrangian L is the
820   same as in Eq. (B.7), and setting $\partial L/\partial B_{ij} = 0$ leads to

$$-\log B_{ij} + \gamma_i - \gamma_j = 0, \quad i \neq j. \tag{B.22}$$

An interchange of the indices and summing the two equations gives

$$\log B_{ij} + \log B_{ji} = 0. \tag{B.23}$$

Hence, $B_{ij} B_{ji} = 1$. A good guess gives $B_{ij}^2 = x_i^*/x_j^*$ and $\gamma_j = \frac{1}{2} \log x_j^*$. From $\frac{\partial}{\partial z_j} L = 0$ we get

$$-\log z_j - \gamma_j = 0, \quad j \in S, \tag{B.24}$$

and in turn $z_j = (x_j^*)^{-1/2}$. Maximality and uniqueness of this solution follow from the strict concavity of $\mathbb{H}(X)$ as a function of B_{ij} and z_j for fixed $\mathbf{x}^*$. We can see this strict concavity by

$$\frac{\partial^2}{\partial B_{ij}^2} \mathbb{H}(X) = \frac{\partial}{\partial B_{ij}} \frac{-x_j^*}{\|\mathbf{u}\|} \log B_{ij} = -\frac{x_j^*}{\|\mathbf{u}\| B_{ij}} < 0 \tag{B.25}$$

and

$$\frac{\partial^2}{\partial z_j^2} \mathbb{H}(X) = \frac{\partial}{\partial z_j} \frac{-x_j^*}{\|\mathbf{u}\|} \log z_j = -\frac{x_j^*}{\|\mathbf{u}\| z_i} < 0. \tag{B.26}$$