Mathematical Geosciences manuscript No.
(will be inserted by the editor)

Information content and maximum entropy of compartmental systems in equilibrium

Holger Metzler · Carlos A. Sierra

Received: Accepted:

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 of ~~lacking~~ the 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 Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgränd 17, 901 83 Umeå, Sweden. In between at: Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Ulls ~~väg~~ 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~~ 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 · MaxEnt · 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.

Code Repository: The Python code to reproduce the figures used in the manuscript is currently provided at
`https://github.com/goujou/entropy_and_complexity_in_eq`.
In case of publication it will be transformed into a permanent repository with a DOI attached to it.

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 (Burnham and Anderson 2002 Höge et al. 2018) . 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 (Golan and Harte 2022) . 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 (Jost 2005) . These metrics have been used extensively to characterize chaotic behavior in complex nonlinear systems ~~, but~~ (Fan et al. 2021), but as we will see later, 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 of material content in compartmental systems 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. 2017 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)~~ (Pressé et al. 2013 Golan and Harte 2022).

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)~~(Trucco 1956 Morzy et al. 2017)

. Thus, they concentrate ~~on only~~ only on 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~~ interpret deterministic compartmental systems from a probabilistic viewpoint. 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), we introduce three entropy measures. 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 focus on a single particle makes our entropies microscopic system properties and consequently distinguishes our approach from the theory of maximum caliber (MaxCal, Jaynes 1985 Roach 2020), where path entropy is interpreted as a macroscopic system property of bulk material. Furthermore, our information theoretical approach differs from the thermodynamic approach to entropy, which has been developed by other authors studying energy transfers and reversibility in thermodynamic systems (Aoki 1988 Haddad et al. 2010 Haddad 2013; 2019).

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~~ provide the fundamentals from information theory and dynamical systems theory that are necessary to introduce the path entropy as the uncertainty of a single particle traveling through the system. ~~Based on this Markov chain we then define three entropy quantities for compartmental systems in equilibrium and adopt the~~

~~MaxEnt theory~~Then, we mathematically derive the path entropy and introduce the entropy rates per unit time and per jump, before we introduce MaxEnt and structural model identification. 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. Then we discuss the results and draw final conclusions.

2 ~~Background and theoretical derivation of~~ Mathematical background: information entropy ~~measures~~and compartmental systems as Markov chains

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

~~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 particlejumps, 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~~”~~.~~

2.1 Short summary of Shannon information entropy

~~Each jump comes with uncertainties about which compartment will be next and how long will the particlestay there. The~~ “~~entropy rate per jump~~” ~~measures the average of these uncertainties with respect to the mean number of jumps until~~ We introduce a few basic concepts of information entropy. Within the framework of this manuscript, discrete entropies are usually associated to a particle's jump into another compartment and differential entropies to a particle's sojourn time within a specific compartment. Entropy rates are defined as average uncertainties of the particle's ~~exit from~~ path while it is in the system.

2.2 ~~Short summary of Shannon information entropy~~

~~We provide a short introduction of basic concepts of information entropy.For~~ See Sects. 2 and 8 of Cover and Thomas (2006) for a more detailed introduction ~~along the lines of Cover and Thomas (2006) see Appendix ??~~to Shannon's information entropy and differential entropy. Entropy rates for discrete- and continuous-time stochastic processes are introduced in Cover and Thomas (Sect. 4 2006) and Bad Dumitrescu (1988)
.

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.

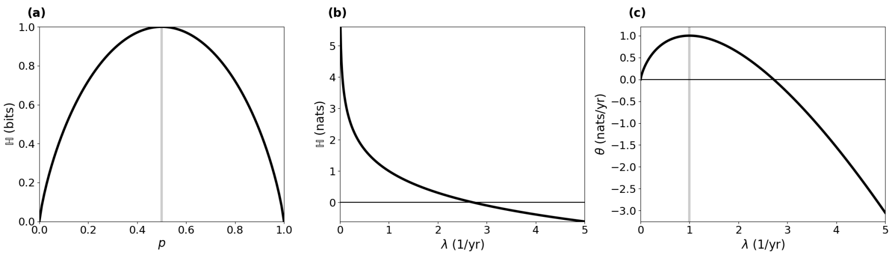


Fig. 1 Shannon entropy of a Bernoulli distribution (a), differential entropy of an exponential distribution (b), and entropy rate of a Poisson process (c). Vertical gray lines indicate the parameter values leading to the highest entropy

~~We can further consider the~~ The "joint entropy" $\cancel{\mathbb{H}(Y_1,Y_2)}$ of two random variables Y_1 and Y_2 ~~, which~~ can be described as

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

where the "conditional entropy" $\cancel{H(Y_2\,|\,Y_1)}$ $\mathbb{H}(Y_2\,|\,Y_1)$ describes the uncertainty of Y_2 under the condition that Y_1 is known. ~~Note that the joint entropy is symmetric, that is,~~ $\cancel{\mathbb{H}(Y_1,Y_2) = \mathbb{H}(Y_2,Y_1)}$ ~~and~~

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

~~with equality if~~ $\cancel{Y_1}$ ~~and~~ $\cancel{Y_2}$ ~~are independent. Consequently,~~

$$\mathbb{H}(Y_2\,|\,Y_1) \leq \mathbb{H}(Y_2)$$

~~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,

$$\theta(Z) = \theta(\text{Poi}(\lambda)) = \lambda\,(1 - \log\lambda). \tag{5}$$

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~~ An autonomous compartmental systems can be described by the d-dimensional linear ODE system

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{x}(t) = \mathrm{B}\,\mathbf{x}(t) + \mathbf{u}, \quad t > 0, \tag{6}$$

with some ~~initial condition $\mathbf{x}(0) = \mathbf{x}^0 \in \mathbb{R}^d$~~ nonnegative 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 $\cancel{\mathbf{u}}$ $\mathbf{u} = (u_i)_{i=1,2,\ldots,d} \in \mathbb{R}^d_+$ is the vector of external inputs to the compartments, and the compartmental matrix $\mathrm{B} \in \mathbb{R}^{d\times d}$ 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 (6) to have reached its unique steady-state or equilibrium compartment vector $\mathbf{x}^* = -\mathrm{B}^{-1}\,\mathbf{u}$. This implies $\|\mathbf{r}\| = \|\mathbf{u}\|$, where $\mathbf{r} = (r_i)_{i=1,2,\ldots,d}$ given by $r_j = z_j\,x_j^*$ is the external outflux vector from the system, and $\|\cdot\|$ denotes the sum of absolute values of a vector (l_1-norm). An open compartmental system in equilibrium given by Eq. (6) is fully characterized by $\mathbf{u}$ and B, and we denote it by $\cancel{M = M(\mathbf{u},\mathrm{B})}$ $M := M(\mathbf{u},\mathrm{B})$.

2.3 The one-particle perspective

While Eq. (6) 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 $\cancel{X_t \in \mathscr{S} := \{1,2,\ldots,d\}}$ $X_t \in S := \{1,2,\ldots,d\}$ 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 $\cancel{\widetilde{\mathscr{S}} := \mathscr{S} \cup \{d+1\}}$ $\widetilde{S} := S \cup \{d+1\}$. Its initial distribution is given by $\widetilde{\beta} = (\beta_1, \beta_2, \ldots, \beta_d, 0)^T$, where $\beta := \mathbf{u}/\|\mathbf{u}\|$, and hence $\beta_j = \mathbb{P}(X_0 = j)$ is the probability of the single particle to enter the system through compartment j. The superscript

T denotes the transpose of the ~~vector/matrixand $\|\mathbf{u}\| = \sum_{i=1}^{d} |u_i|$ denotes the l_1-norm of the vector $\mathbf{u}$. The state-transition~~ respective vector or matrix. The transition-rate matrix of X is given by

$$\mathrm{Q} = \begin{pmatrix} \mathrm{B} & \mathbf{0} \\ \mathbf{z}^T & 0 \end{pmatrix}, \tag{7}$$

and thus

$$\mathbb{P}(X_t = i) = (e^{t\,\mathrm{Q}}\,\widetilde{\beta})_i = \sum_{j=1}^{d} (e^{t\,\mathrm{Q}})_{ij}\,\beta_j, \quad i \in \widetilde{S}, \tag{8}$$

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

$$\mathbb{P}(X_t = i \,|\, X_s = j) = (e^{(t-s)\,\mathrm{Q}})_{ij}, \quad s \le t, \quad i, j \in \widetilde{S}, \tag{9}$$

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. (6) are equivalent, we can write

$$M = M(\mathbf{u}, \mathrm{B}) = M(X). \tag{10}$$

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

$$\mathscr{P}(X) := ((Y_1 = X_0, T_1), (Y_2, T_2), \ldots, (Y_{\mathscr{N}-1}, T_{\mathscr{N}-1}), Y_{\mathscr{N}} = d+1), \tag{11}$$

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

$$\mathscr{N}:=\inf\{n\in\mathbb{N}:Y_n=d+1\} \tag{12}$$

denotes the first hitting time of the absorbing state $d+1$ by the "embedded jump chain" $Y:=(Y_n)_{n=1,2,\dots,\mathscr{N}}$ of X (Norris 1997). With $\lambda_j:=-Q_{jj}$ the one-step transition probabilities of Y are given by, for ~~$i,j\in\widetilde{\mathscr{S}}$~~ $i,j\in\widetilde{S}$,

$$P_{ij}:=\mathbb{P}(Y_{n+1}=i\,|\,Y_n=j)=\begin{cases}0, & i=j \text{ or } \lambda_j=0,\\ Q_{ij}/\lambda_j, & \text{else.}\end{cases} \tag{13}$$

Let $\mathrm{P}|_S=(P_{ij})_{i,j\in S}$ be the restriction of P to S. We can also write ~~$\mathrm{P}=(P_{ij})_{i,j\in\widetilde{\mathscr{S}}}=\mathrm{Q}\,\mathrm{D}^{-1}+\mathrm{I}$, where~~

$$\cancel{\mathrm{D}:=\mathrm{diag}\,(\lambda_1,\lambda_2,\dots,\lambda_d,\lambda_{d+1})}$$

$\mathrm{P}|_S=\mathrm{B}\,\mathrm{D}^{-1}+\mathrm{I}$, where $\mathrm{D}:=\mathrm{diag}\,(\lambda_1,\lambda_2,\dots,\lambda_d)$ is the diagonal matrix with the diagonal entries of ~~Q~~ B, and I denotes the identity matrix of appropriate dimension. ~~We define the matrix $\mathrm{P_B}:=(P_{ij})_{i,j\in\mathscr{S}}$, then $\mathrm{M}:=(\mathrm{I}-\mathrm{P_B})^{-1}$~~ Then $\mathrm{M}:=(\mathrm{I}-\mathrm{P}|_S)^{-1}$ 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 ~~$i\in\mathscr{S}$~~ $i\in S$ is given by

$$\mathbb{E}\,[N_i]=\sum_{j=1}^{d}M_{ij}\,\beta_j=(\mathrm{M}\,\beta)_i=\left[(\mathrm{I}-\mathrm{P}_{\cancel{\mathrm{B}}}|_S)^{-1}\,\beta\right]_i=(\mathrm{D}\,\mathrm{B}^{-1}\,\beta)_i=\frac{\lambda_i\,x_i^*}{\|\mathbf{u}\|} \tag{14}$$

and the total expected number of jumps is given by

$$\mathbb{E}\,[\mathscr{N}]=\sum_{i=1}^{d}(\mathrm{M}\,\beta)_i+1=\sum_{i=1}^{d}\mathbb{E}\,[N_i]+1, \tag{15}$$

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

$$\mathscr{T} := \inf\{t > 0 : X_t = d+1\}. \tag{16}$$

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

$$f_{\mathscr{T}}(t) = \mathbf{z}^T\, e^{t\,\mathrm{B}}\,\beta, \quad t \geq 0. \tag{17}$$

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. 2017) of system (6) in equilibrium and is given by total stocks over total fluxes, that is,

$$\mathbb{E}\,[\mathscr{T}] = \frac{\|\mathbf{x}^*\|}{\|\mathbf{u}\|}. \tag{18}$$

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

$$\mathbb{1}_{\{A\}} = \begin{cases} 1, & A \text{ is true,} \\ 0, & \text{else,} \end{cases} \tag{19}$$

then $O_j := \sum_{k=1}^{\mathscr{N}-1} \mathbb{1}_{\{Y_k=j\}}\, T_k$ 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)

$$\mathbb{E}\,[O_j] = \frac{x_j^*}{\|\mathbf{u}\|}, \tag{20}$$

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

3 Entropy measures, MaxEnt, and structural model identification

Based on these basic structures of the path of a single particle traveling through the system, we compute three different types of entropy, for which we provide below a summary of the desirable relations among them:

(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.

Once these entropy metrics are established, we introduce MaxEnt and show how to apply it to the problem of structural model identification.

3.1 Path entropy, entropy rate per unit time, and entropy rate per jump

The path $\mathscr{P} = \mathscr{P}(X)$ given by Eq. (11) 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~~ σ-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

$$
\begin{aligned}
f_{\mathscr{P}}(p) = \beta_{x_1} \left(\prod_{j=1}^{d} \prod_{i=1, i \neq j}^{d+1} (Q_{ij})^{N_{ij}(p)} \right) \prod_{j=1}^{d} e^{-\lambda_j R_j(p)}, \\
p = ((x_1, t_1), (x_2, t_2), \dots, (x_n, t_n), d+1) \in \wp.
\end{aligned} \tag{21}
$$

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

$$
\begin{aligned}
&\mathbb{P}((Y_1 = x_1, T_1 \le t_1), \dots, (Y_n = x_n, T_n \le t_n), Y_{n+1} = d+1) \\
&\quad = \mathbb{P}(Y_{n+1} = d+1 \,|\, Y_n = x_n) \\
&\qquad \cdot \prod_{k=2}^{n} \mathbb{P}(Y_k = x_k, T_k \le t_k \,|\, Y_{k-1} = x_{k-1}) \, \mathbb{P}(Y_1 = x_k, T_1 \le t_1) \\
&\quad = P_{d+1, x_n} \left[\prod_{k=2}^{n} P_{x_k x_{k-1}} \left(1 - e^{-\lambda_{x_k} t_k}\right) \right] \beta_{x_1} \left(1 - e^{-\lambda_{x_1} t_1}\right) \\
&\quad = \int_{\mathbb{T}_n} \beta_{x_1} \prod_{k=1}^{n} Q_{x_{k+1} x_k} \, e^{-\lambda_{x_k} \tau_k} \, \mathrm{d}\tau_1 \mathrm{d}\tau_2 \cdots \mathrm{d}\tau_n
\end{aligned} \tag{22}
$$

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

$$
\begin{aligned}
f_{\mathscr{P}}(p) = \beta_{x_1} \prod_{k=1}^{n} Q_{x_{k+1} x_k} \, e^{-\lambda_{x_k} t_k}, \\
p = ((x_1, t_1), (x_2, t_2), \dots, (x_n, t_n), d+1) \in \wp.
\end{aligned} \tag{23}
$$

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

$$
\begin{aligned}
\prod_{k=1}^{n} e^{-\lambda_{x_k} t_k} &= \prod_{k=1}^{n} \prod_{j=1}^{d} \mathbb{1}_{\{x_k = j\}} \, e^{-\lambda_j t_k} = \prod_{j=1}^{d} e^{-\lambda_j \sum_{k=1}^{n} \mathbb{1}_{\{x_k = j\}} t_k} \\
&= \prod_{j=1}^{d} e^{-\lambda_j R_j(p)}.
\end{aligned} \tag{24}
$$

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 $[0, \mathscr{T}]$, 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

$$\begin{aligned}\mathbb{H}(X) &= \mathbb{H}(\mathscr{P}) \\ &= -\sum_{i=1}^{d} \beta_i \log \beta_i \\ &\quad + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i\neq j}^{d} B_{ij}\,(1-\log B_{ij}) + z_j\,(1-\log z_j) \right].\end{aligned} \tag{25}$$

Proof Let X have the finite path representation

$$\mathscr{P} = \mathscr{P}(X) = ((Y_1, T_1), (Y_2, T_2), \ldots, (Y_n, T_n), d+1) \tag{26}$$

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

$$-\log f_{\mathscr{P}}(\mathscr{P}) = -\log \beta_{Y_1} - \sum_{j=1}^{d} \sum_{i=1, i\neq j}^{d+1} N_{ij}(\mathscr{P}) \log Q_{ij} + \sum_{j=1}^{d} \lambda_j R_j(\mathscr{P}). \tag{27}$$

We compute the expectation and get

$$\begin{aligned}\mathbb{H}(X) = \mathbb{H}(\mathscr{P}) &= -\mathbb{E}\left[\log f_{\mathscr{P}}(\mathscr{P})\right] \\ &= -\mathbb{E}\left[\log \beta_{Y_1}\right] - \sum_{j=1}^{d} \sum_{i=1, i\neq j}^{d+1} \mathbb{E}\left[N_{ij}(\mathscr{P})\right] \log Q_{ij} + \sum_{j=1}^{d} \lambda_j\, \mathbb{E}\left[R_j(\mathscr{P})\right] \\ &= \mathbb{H}(Y_1) + \sum_{j=1}^{d} \lambda_j\, \mathbb{E}\left[R_j(\mathscr{P})\right] - \sum_{j=1}^{d} \sum_{i=1, i\neq j}^{d+1} \mathbb{E}\left[N_{ij}(\mathscr{P})\right] \log Q_{ij}.\end{aligned} \tag{28}$$

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. (14)

and (13),

$$\mathbb{E}\left[N_{ij}(\mathscr{P})\right] = \mathbb{E}\left[N_j(\mathscr{P})\right] P_{ij} = \begin{cases} \frac{x_j^*}{\|\mathbf{u}\|} B_{ij}, & i \le d, \\ \frac{x_j^*}{\|\mathbf{u}\|} z_j, & i = d+1. \end{cases} \tag{29}$$

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

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(Y_1) + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\left(\sum_{i=1, i\neq j}^{d} B_{ij} + z_j \right) \right. \\ &\qquad \left. - \sum_{i=1, i\neq j}^{d} B_{ij} \log B_{ij} - z_j \log z_j \right] \\ &= - \sum_{i=1}^{d} \beta_i \log \beta_i + \sum_{j=1}^{d} \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i\neq j}^{d} B_{ij} \left(1 - \log B_{ij}\right) \right. \\ &\qquad \left. + z_j \left(1 - \log z_j\right) \right]. \end{aligned} \tag{30}$$

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

$$\mathbb{H}(\beta) := - \sum_{i=1}^{d} \beta_i \log \beta_i. \tag{31}$$

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

$$\mathbb{H}(X) = \mathbb{H}(\beta) + \sum_{j=1}^{d} \mathbb{E}\left[O_j\right] \left(\sum_{i=1, i\neq j}^{d} \theta(\mathrm{Poi}(B_{ij})) + \theta(\mathrm{Poi}(z_j)) \right) \tag{32}$$

and

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\beta) \\ &\quad + \sum_{j=1}^{d} \mathbb{E}\left[N_j\right] \left(\mathbb{H}(\mathrm{Exp}(\lambda_j)) + \mathbb{H}(P_{1,j}, P_{2,j}, \ldots, P_{d,j}, P_{d+1,j}) \right), \end{aligned} \tag{33}$$

which can be rewritten as

$$\mathbb{H}(X) = \mathbb{H}(\beta) + \sum_{j=1}^{d} \mathbb{E}[N_j]\, \mathbb{H}(P_{1,j}, P_{2,j}, \ldots, P_{d,j}, P_{d+1,j}) \tag{34}$$

$$+ \sum_{j=1}^{d} \mathbb{E}[N_j]\, \mathbb{H}(\mathrm{Exp}(\lambda_j)). \tag{35}$$

Proof By virtue of Eq. (32), we replace $x_j^*/\|\mathbf{u}\|$ by $\mathbb{E}[O_j]$ in Eq. (25) and take into account that the entropy rate of a Poisson process with intensity rate λ equals $\lambda\,(1-\log\lambda)$ to prove Eq. (32). To prove Eq. (33) we use Eq. (14) to replace $x_j^*/\|\mathbf{u}\|$ in Eq. (25) by $\mathbb{E}[N_j]/\lambda_j$ and obtain

$$\begin{aligned}\mathbb{H}(X) = &-\sum_{i=1}^{d} \beta_i \log \beta_i + \sum_{j=1}^{d} \mathbb{E}[N_j]\,(1-\log\lambda_j) \\ &+ \sum_{j=1}^{d} \mathbb{E}[N_j] \left(-\sum_{i=1, i\neq j}^{d} \frac{B_{ij}}{\lambda_j} \log \frac{B_{ij}}{\lambda_j} - \frac{z_j}{\lambda_j} \log \frac{z_j}{\lambda_j} \right).\end{aligned} \tag{36}$$

Here, $(1-\log\lambda_j)$ is the entropy of an exponential random variable with rate parameter λ_j. Using definition (13) of P_{ij} we replace B_{ij}/λ_j by P_{ij} for ~~$i \in \mathscr{S}$~~ $i \in S$ and z_j/λ_j by $P_{d+1,j}$ and finish the proof.

~~We now define the~~ "~~path entropy~~" ~~of the compartmental system~~ By identifying a compartmental system $M = M(\mathbf{u}, \mathrm{B})$ with its associated absorbing continuous-time Markov chain X and the according path $\mathscr{P} = \mathscr{P}(X)$ of a single traveling particle, we transfer the concept of the path entropy $\mathbb{H}(\mathscr{P})$ from the probabilistic to the deterministic realm.

Definition 1 The "path entropy of the compartmental system" M in equilibrium ~~,~~ given by Eq. (6) ~~, as the path entropy of its associated~~ with associated absorbing continuous-time Markov chain X ~~, that is~~and path $\mathscr{P} = \mathscr{P}(X)$, is defined by the path entropy

$$\mathbb{H}(M\mathscr{P}) := \mathbb{H}(\mathscr{P}(X)) = \mathbb{H}(\mathscr{P}(X)). \tag{37}$$

~~For~~

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\,x(t) = -\lambda\,x(t) + u, \quad t > 0, \tag{38}$$

~~the~~ and denote its associated path by $\mathscr{P}_\lambda$. The entropy of the initial distribution vanishes, and we obtain

$$\mathbb{H}(\cancel{M}\mathscr{P}_\lambda) = \frac{x^*}{u}\,\lambda\,(1-\log\lambda) = \frac{1}{\lambda}\,\lambda\,(1-\log\lambda) = 1-\log\lambda, \tag{39}$$

which equals the differential entropy $1-\log\lambda$ of the exponentially distributed mean transit time $\mathscr{T}_\lambda \sim \mathrm{Exp}(\lambda)$, 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 $\cancel{\mathbb{H}(M_\lambda) = \lambda^{-1}\lambda\,(1-\log\lambda)}$ $\mathbb{H}(\mathscr{P}_\lambda) = \lambda^{-1}\lambda\,(1-\log\lambda)$ as the instantaneous Poisson entropy rate $\lambda\,(1-\log\lambda)$ multiplied with the expected duration $\cancel{\mathbb{E}[\mathscr{T}] = \lambda^{-1}}$ $\mathbb{E}[\mathscr{T}_\lambda] = \lambda^{-1}$ of the particle's stay in the system.

Migrating to a d-dimensional system, we can interpret $\cancel{\mathbb{H}(M)}$ $\mathbb{H}(\mathscr{P})$ as the entropy of a continuous-time process in the light of Eq. (32) and as the entropy of a discrete-time process in the light of Eq. (33). 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, $\theta(\mathrm{Poi}(B_{ij}))$, and out of the system, $\theta(\mathrm{Poi}(z_j))$, where the weights are the expected occupation times of the different compartments $\cancel{j \in \mathscr{S}}$ $j \in S$. 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, $\mathbb{H}(\mathrm{Exp}(\lambda_j))$, and the uncertainty of where to jump afterwards, ~~$\mathbb{H}(\{P_{ij} : i \in \widetilde{\mathscr{S}}, j \in \mathscr{S}, i \neq j\})$~~$\mathbb{H}(\{P_{ij} : i \in \widetilde{S}, j \in S, i \neq j\})$. Hence, in the light of Eq. (33), we can separate the path entropy into a discrete part associated to jump uncertainty given by Eq. (34) and a continuous part associated to sojourn time uncertainty given by Eq. (35).

The two interpretations of the path entropy ~~$\mathbb{H}(M)$~~ $\mathbb{H}(\mathscr{P})$ (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

$$\theta(M\mathscr{P}) = \frac{\mathbb{H}(M)}{\mathbb{E}[\mathscr{T}]}\frac{\mathbb{H}(\mathscr{P})}{\mathbb{E}[\mathscr{T}]} \tag{40}$$

and the "entropy rate per jump" by

$$\theta_J(M\mathscr{P}) = \frac{\mathbb{H}(M)}{\mathbb{E}[\mathscr{N}]}\frac{\mathbb{H}(\mathscr{P})}{\mathbb{E}[\mathscr{N}]}. \tag{41}$$

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. We can see that by considering the the stationary process $Z = (Z_n)_{n\geq 1} = (\widetilde{Y}_n, \widetilde{T}_n)_{n\geq 1}$ on the space $(\widetilde{S} \times \mathbb{R}_+)$ defined by the transition probabilities $\widetilde{P}_{ij}(t) = \mathbb{P}(\widetilde{Y}_{n+1} = i, \widetilde{T}_{n+1} \leq t \,|\, \widetilde{Y}_n = j)$

given by

$$\widetilde{P}_{ij}(t) = \begin{cases} 0, & i = j, \\ B_{ij}\,\lambda_j^{-1}\,(1 - e^{-\lambda_i t}), & i, j \le d,\, i \ne j, \\ z_j\,\lambda_j^{-1}, & i = d+1,\, j \le d, \\ \beta_i\,(1 - e^{-\lambda_i t}), & i \le d,\, j = d+1, \end{cases} \tag{42}$$

and initial (stationary) distribution

$$\pi_j(t) = \frac{1}{\mathbb{E}\,[\mathscr{N}]} \cdot \begin{cases} \mathbb{E}\,[N_j]\,(1 - e^{-\lambda_j t}), & j \le d, \\ 1, & j = d+1. \end{cases} \tag{43}$$

This process describes the infinite journey, that is the sequence of visited compartments with the associated sojourn times, of a single particle through the system with immediate jumps back into the system when leaving it.

Proposition 2 The entropy rate per jump, $\theta_J(\mathscr{P})$, equals the entropy rate of the stationary process Z.

Proof Step 1. We show that $Z = (\widetilde{Y}, \widetilde{T})$ is stationary. To that end, we define $\pi_j := \lim_{t\to\infty} \pi_j(t)$, and we prove $\mathbb{P}(\widetilde{Y}_2 = i, \widetilde{T}_2 \le t) = \pi_i(t) = \mathbb{P}(\widetilde{Y}_1 = i, \widetilde{T}_1 \le t)$. Stationarity follows then by induction. Let $i = d+1$. Then,

$$\begin{aligned} \mathbb{P}(\widetilde{Y}_2 = i, \widetilde{T}_2 \le t) &= \sum_{j=1}^{d} \mathbb{P}(\widetilde{Y}_2 = i, \widetilde{T}_2 \le t \,|\, \widetilde{Y}_1 = j)\,\mathbb{P}(\widetilde{Y}_1 = j) \\ &= \sum_{j=1}^{d} \widetilde{P}_{d+1,j}(t)\,\pi_j \\ &= \sum_{j=1}^{d} \frac{z_j}{\lambda_j}\,\frac{\mathbb{E}\,[N_j]}{\mathbb{E}\,[\mathscr{N}]}. \end{aligned} \tag{44}$$

415 By Eq. (14), $r_j = z_j x_j^*$, and $\|\mathbf{r}\| = \|\mathbf{u}\|$, we get

$$\mathbb{P}(\widetilde{Y}_2 = i, \widetilde{T}_2 \le t) = \frac{1}{\mathbb{E}[\mathscr{N}]} \sum_{j=1}^{d} \frac{z_j}{\lambda_j} \frac{\lambda_j x_j^*}{\|\mathbf{u}\|} = \frac{1}{\mathbb{E}[\mathscr{N}]} \frac{\mathbf{z}^T \mathbf{x}^*}{\|\mathbf{u}\|} = \pi_{d+1}(t). \tag{45}$$

416 Now let $i \le d$. Then

$$\begin{aligned}
\mathbb{P}(\widetilde{Y}_2 = i, \widetilde{T}_2 \le t) &= \sum_{j=1, j\ne i}^{d} \frac{B_{ij}}{\lambda_j} (1 - e^{-\lambda_i t}) \frac{\mathbb{E}[N_j]}{\mathbb{E}[\mathscr{N}]} + \beta_i (1 - e^{-\lambda_i t}) \frac{1}{\mathbb{E}[\mathscr{N}]} \\
&= \frac{1}{\mathbb{E}[\mathscr{N}]} \left[\sum_{i=1, i\ne j}^{d} \frac{B_{ij} x_j^*}{\|\mathbf{u}\|} + \beta_i \right] (1 - e^{-\lambda_i t}) \\
&= \frac{1}{\mathbb{E}[\mathscr{N}]} \left[\frac{(\mathrm{B}\,\mathbf{x}^*)_i}{\|\mathbf{u}\|} - \frac{B_{ii} x_i^*}{\|\mathbf{u}\|} + \beta_i \right] (1 - e^{-\lambda_i t}) \\
&= \frac{1}{\mathbb{E}[\mathscr{N}]} \left[-\frac{u_i}{\|\mathbf{u}\|} + \frac{\lambda_i x_i^*}{\|\mathbf{u}\|} + \beta_i \right] (1 - e^{\lambda_i t}) \\
&= \frac{1}{\mathbb{E}[\mathscr{N}]} \mathbb{E}[N_i] (1 - e^{-\lambda_i t}) \\
&= \pi_i(t).
\end{aligned} \tag{46}$$

417 Step 2. Since Z is stationary, by Cover and Thomas (2006, Theorem 4.2.1), its
418 entropy rate given by

$$\theta(Z) = \lim_{n\to\infty} \mathbb{H}(Z_{n+1} \,|\, Z_n) = \mathbb{H}(Z_2 \,|\, Z_1), \tag{47}$$

419 which computes to

$$\begin{aligned}
\theta(Z) &= \mathbb{H}((\widetilde{Y}_2, \widetilde{T}_2) \,|\, (\widetilde{Y}_1, \widetilde{T}_1)) = \mathbb{H}((\widetilde{Y}_2, \widetilde{T}_2) \,|\, \widetilde{Y}_1) = \mathbb{H}(\widetilde{T}_2 \,|\, \widetilde{Y}_2, \widetilde{Y}_1) + \mathbb{H}(\widetilde{Y}_2 \,|\, \widetilde{Y}_1) \\
&= \mathbb{H}(\widetilde{T}_2 \,|\, \widetilde{Y}_2) + \mathbb{H}(\widetilde{Y}_2 \,|\, \widetilde{Y}_1).
\end{aligned} \tag{48}$$

By stationarity, $\mathbb{H}(\widetilde{T}_2\,|\,\widetilde{Y}_2) = \mathbb{H}(\widetilde{T}_1\,|\,\widetilde{Y}_1)$. Consequently,

$$\begin{aligned}\theta(Z) &= \mathbb{H}(\widetilde{T}_1\,|\,\widetilde{Y}_1) + \mathbb{H}(\widetilde{Y}_2\,|\,\widetilde{Y}_1)\\ &= \sum_{j=1}^{d+1} \pi_j \left[\mathbb{H}(\widetilde{T}_1\,|\,\widetilde{Y}_1 = j) + \mathbb{H}(\widetilde{Y}_2\,|\,\widetilde{Y}_1 = j)\right],\\ &= \frac{1}{\mathbb{E}\,[\mathscr{N}]}\left(\sum_{j=1}^{d} \mathbb{E}\,[N_j]\left[\mathbb{H}(\widetilde{T}_1\,|\,\widetilde{Y}_1 = j) + \mathbb{H}(\widetilde{Y}_2\,|\,\widetilde{Y}_1 = j)\right] + \mathbb{H}(\widetilde{Y}_2\,|\,\widetilde{Y}_1 = d+1)\right),\end{aligned} \tag{49}$$

which together with Eq. (33) finishes the proof.

If we divide the entropy rate per jump by the average time between two jumps, we obtain the entropy rate per unit time. The average time between two jumps is given by

$$\sum_{j=1}^{d} \pi_j\,\lambda_j^{-1} = \frac{1}{\mathbb{E}\,[\mathscr{N}]}\sum_{j=1}^{d}\frac{x_j^*}{\|\mathbf{u}\|} = \frac{\mathbb{E}\,[\mathscr{T}]}{\mathbb{E}\,[\mathscr{N}]}. \tag{50}$$

Hence,

$$\theta(\mathscr{P}) = \frac{\mathbb{E}\,[\mathscr{N}]}{\mathbb{E}\,[\mathscr{T}]}\,\theta(Z) \tag{51}$$

is the average uncertainty per unit time of the stationary process Z.

3.2 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. In the next two examples we identify compartmental models with maximum entropy under some restrictions. Both examples show that maximizing entropy means also maximizing symmetry, as much as the given constraints allow.

Example 1 Consider the set $\mathscr{M}_1$ of equilibrium compartmental systems (6) 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 $M_1^* \in \mathscr{M}_1$ with path $\mathscr{P}_1^* := \mathscr{P}(M_1^*)$ such that the ~~entropy rate per unit time $\theta(M_1^*)$~~path entropy $\mathbb{H}(\mathscr{P}_1^*)$, or equivalently~~the path entropy $\mathbb{H}(\mathscr{P}(M_1^*))$,~~, the entropy rate per unit time $\theta(\mathscr{P}_1^*)$ is maximized. We can show (see Proposition A.1) that the compartmental system $M_1^* = M(\mathbf{u}, \mathrm{B})$ with

$$\mathrm{B} = \begin{pmatrix} -\lambda & 1 & \cdots & 1 \\ 1 & -\lambda & 1\cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & -\lambda \end{pmatrix}, \tag{52}$$

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~~ subclass $\mathscr{M}_2 \subseteq \mathscr{M}_1$ of compartmental models from the previous example with the additional restriction of a predefined positive steady-state vector $\mathbf{x}^*$. Then the compartmental system $M_2^* = M(\mathbf{u},\mathrm{B})$ with path $\mathscr{P}_2^*$ and

$$B_{ij} = \begin{cases} \sqrt{\frac{x_i^*}{x_j^*}}, & i \neq j, \\ -\sum\limits_{k=1,k\neq j}^{d} \sqrt{\frac{x_k^*}{x_j^*}} - \frac{1}{\sqrt{x_j^*}}, & i = j, \end{cases} \tag{53}$$

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

3.3 Structural model identification ~~via~~ assisted by MaxEnt

Suppose we observe a natural system and conduct measurements from which we try to construct a linear autonomous compartmental model in equilibrium that represents the observed natural system as well as possible. The first question that arises is the one for the number of compartments the model should ideally have. MaxEnt cannot be helpful here because by adding more and more compartments we can theoretically increase the entropy of the model indefinitely. Consequently, the problem of finding the right dimension of system (6) has to be solved by other means. One way to do this is to analyze an impulse response function of the system and its Laplace transform, that is the transfer function of the system, and identify the most dominating frequencies. The impulse response or the transfer function might be possible to obtain by tracer experiments (Anderson 1983 Walter 1986).

Once the desired number of compartments is identified, we can focus on the structure and values of external input and output fluxes as well as internal fluxes. In Anderson (1983, Chapter 16) the “structural identification problem” of linear autonomous systems is described as follows. Suppose we are interested in determining a d-dimensional system of form (6). We are interested in sending an impulse into the system at time $t = 0$ and analyzing its further behavior. To that end, we rewrite the system to

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbf{x}(t) &= \mathrm{B}\,\mathbf{x}(t) + \mathrm{A}\,\mathbf{u}, \quad t \geq 0,\\ \mathbf{x}(0) &= \mathbf{0} \in \mathbb{R}^d,\\ \mathbf{y}(t) &= \mathrm{C}\,\mathbf{x}(t), \quad t \geq 0.\end{aligned} \tag{54}$$

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

$$\begin{aligned}\mathbf{y}(t) = \mathrm{C}\,\mathbf{x}(t) &= \mathrm{C}\int_0^t e^{(t-\tau)\,\mathrm{B}}\,\mathrm{A}\,\mathbf{u}(\tau)\,\mathrm{d}\tau\\ &= \left[\mathrm{C}\,e^{t\,\mathrm{B}}\,\mathrm{A}\right] * \mathbf{u}(t),\end{aligned}$$

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

$$\Psi(t) := \mathrm{C}\,e^{t\,\mathrm{B}}\,\mathrm{A}, \quad t \geq 0, \tag{55}$$

or in the "transfer function"

$$\widehat{\Psi}(s) := \mathrm{C}\,(s\,\mathrm{I} - \mathrm{B})^{-1}\,\mathrm{A}, \quad s \geq 0, \tag{56}$$

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 (54) "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 “~~nonidentifiable~~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.

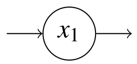
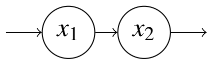
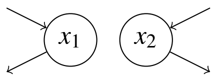
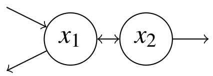
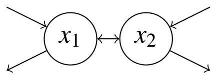
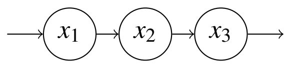
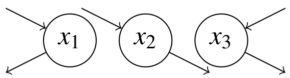
4 Application to particular systems

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~~concepts. Then we compute entropy quantities for two carbon-cycle models in dependence on environmental and biochemical parameters. ~~Afterwards,~~ At last we apply MaxEnt to solve an equifinality problem in model selection as an example for how to tackle this problem arising from, for instance, tracer experiments.

4.1 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.

Table 1 Overview of different entropy measures of simple models with different structures. The columns from left to right represent a schematic representation of the model structure, its mathematical representation, entropy rate per jump θ_J, mean number of jumps $\mathbb{E}[\mathscr{N}]$, entropy rate per unit time θ, mean transit time $\mathbb{E}[\mathscr{T}]$, and path entropy $\mathbb{H}(\mathscr{P})$. Underlined numbers are the highest values per column

Structure	$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t)$	θ_J	$\mathbb{E}[\mathscr{N}]$	θ	$\mathbb{E}[\mathscr{T}]$	$\mathbb{H}(\mathscr{P})$
	$-\lambda x + 1$	$0.5\,(1-\log\lambda)$	2.00	$\lambda\,(1-\log\lambda)$	$1/\lambda$	$1-\log$
	$\begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	0.67	3.00	1.00	2.00	2.00
	$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	0.85	2.00	1.69	1.00	1.69
	$\begin{pmatrix} -1 & 1/2 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	1.08	5.00	1.35	$\underline{4.00}$	$\underline{5.39}$
	$\begin{pmatrix} -1 & 1/2 \\ 1/2 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	$\underline{1.36}$	3.00	2.04	2.00	4.08
	$\begin{pmatrix} -1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	0.75	4.00	1.00	3.00	3.00
	$\begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$	1.05	2.00	$\underline{2.10}$	1.00	2.10

4.2 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 $x_1 = 37\,\mathrm{PgC}$,

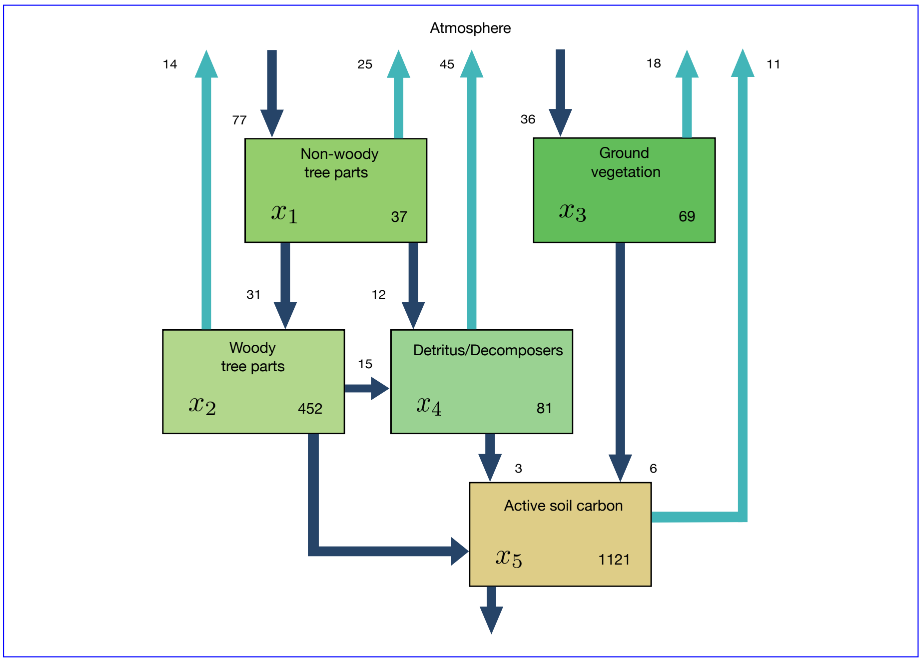


Fig. 2 Schematic of the linear autonomous global carbon cycle model in steady state introduced by Emanuel et al. (1981)

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 potentially increase and speed up the system with increasing global surface temperature (Sierra et al. 2023). For a given ξ the equilibrium model $M_\xi = M(\mathbf{u}, \mathrm{B}_\xi)$ is given by

$$\mathbf{u} = (77;\, 0;\, 36;\, 0;\, 0)^T\,\mathrm{PgC\,yr^{-1}} \tag{57}$$

and

$$\mathrm{B}_\xi = \xi \begin{pmatrix} -77/37 & 0 & 0 & 0 & 0 \\ 31/37 & -31/452 & 0 & 0 & 0 \\ 0 & 0 & -36/69 & 0 & 0 \\ 21/37 & 15/452 & 12/69 & -48/81 & 0 \\ 0 & 2/452 & 6/69 & 3/81 & -11/1,121 \end{pmatrix} \mathrm{yr}^{-1}, \tag{58}$$

where the numbers are chosen as in Thompson and Randerson (1999). The input vector is expressed in units of petagrams of carbon per year ($\mathrm{PgC\,yr^{-1}}$) 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. (34)). The solid line marked by circles represents the part of the path entropy related to sojourn-associated uncertainties (Eq. (35)), which as a weighted average of one-pool entropies decreases ~~similar to~~ similarly as 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 decreasing uncertainty associated to sojourn times ~~more and more~~ increasingly 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 process with intensity rate $\lambda > 1$ (see Fig. 1, panel (c)). The entropy rate per jump (panel ~~f~~(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~~ this multiple-pool 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.

4.3 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-

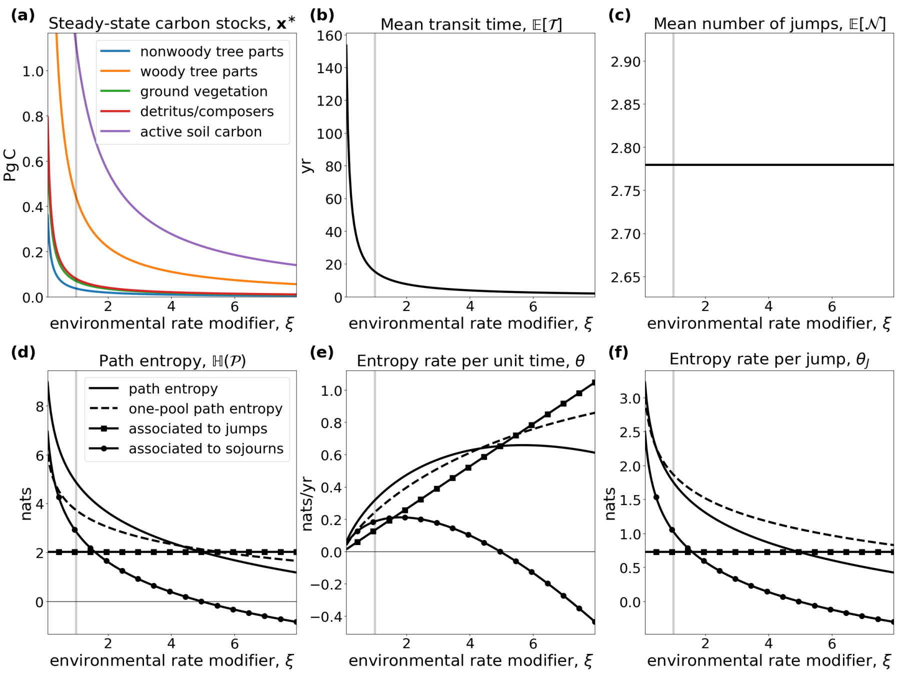


Fig. 3 (a) Equilibrium carbon stocks and (b)–(f) entropy related quantities of the global carbon cycle model introduced by Emanuel et al. (1981) in dependence on the environmental rate coefficient ξ

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

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} C_s \\ C_b \end{pmatrix}(t) = \begin{pmatrix} -\lambda(\mathbf{x}(t)) & \mu_b \\ \varepsilon\lambda(\mathbf{x}(t)) & -\mu_b \end{pmatrix}\begin{pmatrix} C_s \\ C_b \end{pmatrix} + \begin{pmatrix} F_{\mathrm{NPP}} \\ 0 \end{pmatrix}, \tag{59}$$

where $\mathbf{x}(t) = (C_s, C_b)^T(t)$. We denote by C_s and C_b ~~soil~~ substrate 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 (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 (yr^{-1}) and the half-saturation constant for soil carbon assimilation by microbial biomass ($\mathrm{gC\,m^{-2}}$), respectively.

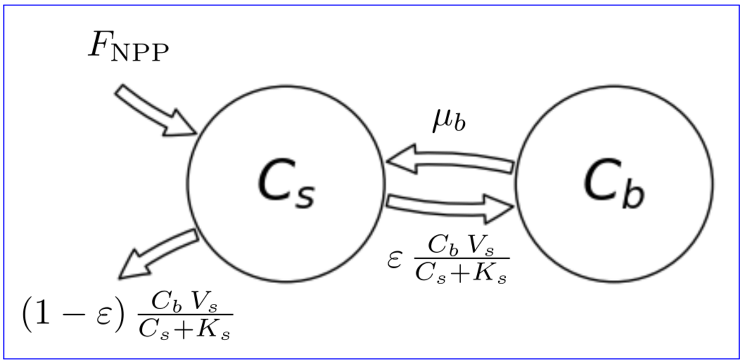


Fig. 4 Scheme of the nonlinear autonomous carbon cycle model introduced by Wang et al. (2014) with two compartments: substrate organic carbon (C_s) and microbial biomass (C_b)

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

$$C_s^* = \frac{K_s}{\frac{V_s\varepsilon}{\mu_b} - 1} \quad \text{and} \quad C_b^* = \frac{F_{\mathrm{NPP}}}{\mu_b\left(-1+\frac{1}{\varepsilon}\right)}. \tag{60}$$

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

$$\lambda(\mathbf{x}) = \frac{C_b V_s}{C_s + K_s}. \tag{61}$$

From Wang et al. (2014) we take the parameter values $\mu_b = 4.38\,\mathrm{yr}^{-1}$, $F_{\mathrm{NPP}} = 345.00\,\mathrm{gC\,m^{-2}\,yr^{-1}}$, 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. (61), 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 $M_\lambda = M(345.00\,\mathrm{g\,C\,m^{-2}\,yr^{-1}}, -1/\mathbb{E}[\mathscr{T}_\varepsilon])$ 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 uncertainties mostly balance out (dashed lines in panels ~~d–f~~(d)–(f)).

4.4 Model identification via Maxent

The following example is inspired by Anderson (1983, Example 16 C). It shows how MaxEnt can help ~~take a decision~~ make a decision about 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

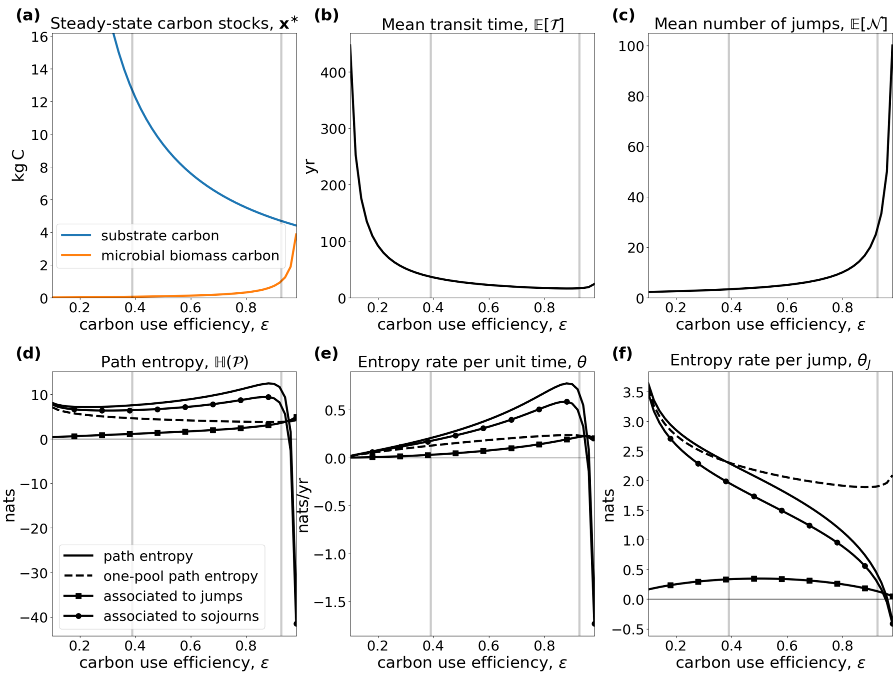


Fig. 5 (a) Equilibrium carbon stocks and (b)–(f) entropy related quantities of the global carbon cycle model introduced by Wang et al. (2014) in dependence on the microbial carbon use efficiency ε

2-dimensional equilibrium compartmental system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}(t) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mathrm{gC\,yr^{-1}}, \quad t > 0. \tag{62}$$

We immediately notice that $\mathbf{u} = (1,0)^T\,\mathrm{gC\,yr^{-1}}$ and $\mathrm{A} = \mathrm{I}$. Further, we decide to measure the contents of compartment 1 such that $\mathrm{C} = (1,0)$. 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

$$\widehat{\Psi}(s) = \frac{s + \gamma_1}{s^2 + \gamma_2\, s + \gamma_3}, \tag{63}$$

where

$$\begin{aligned}
\gamma_1 &= B_{12} + z_2, \\
\gamma_2 &= B_{21} + z_1 + B_{12} + z_2, \\
\gamma_3 &= z_1\,B_{12} + z_1\,z_2 + B_{21}\,z_2.
\end{aligned} \tag{64}$$

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}, and B_{22}, or equivalently, B_{12}, B_{21}, z_1, and z_2, but only three equations to determine them. Consequently, the system is ~~nonidentifiable~~ non-identifiable and there remains a class $\mathscr{M}$ of models which all satisfy Eq. (64). Which model out of $\mathscr{M}$ are we going to select now?

Here, MaxEnt comes into play. We intend to select the model that best represents the information given by our measurement data. We have to find $M^* = M(\mathbf{u}, \mathrm{B}^*)$ such that

$$M^* = \underset{M \in \mathscr{M}}{\arg\max}\; \theta(\mathscr{P}(M)). \tag{65}$$

~~We maximize~~ Maximizing the entropy rate per unit time here ~~instead~~ leads to a feasible optimization problem, whereas maximization of the path entropy ~~, because~~ by slowing down the model ~~we could potentially increase~~ and indefinitely increasing its mean transit time and with it its path entropy ~~indefinitely~~would lead to an unbounded optimization problem. The parameter space associated to $\mathscr{M}$ is given by

$$\{\mathbf{p} = (B_{12}, B_{21}, z_1, z_2) \in \mathbb{R}^4_+ : \mathbf{p} \text{ satisfies Eq. (64)}\}, \tag{66}$$

and is not guaranteed to be convex in general. Consequently, by fundamental principles from mathematical optimization theory, existence and uniqueness of M^* are not guaranteed and we must apply optimization methods tailored to the specific case at hand.

Let us turn to a numerical example in which we suppose to be given $\gamma_1 = 3\,\mathrm{yr}^{-1}$, $\gamma_2 = 5\,\mathrm{yr}^{-1}$, and $\gamma_3 = 4\,\mathrm{yr}^{-1}$. ~~A nonlinear optimization algorithm with the arbitrarily~~

~~chosen initial values $B_{12} = 3\,\mathrm{yr}^{-1}$, $B_{21} = 0\,\mathrm{yr}^{-1}$, $z_1 = 1\,\mathrm{yr}^{-1}$, and $z_2 = 1\,\mathrm{yr}^{-1}$ ends approximately with the terminal compartmental matrix~~

$$\cancel{\mathrm{B}^* \approx \begin{pmatrix} -2.00 & 1.90 \\ 1.05 & -3.00 \end{pmatrix} \mathrm{yr}^{-1}}$$

~~and the terminal~~ Since convexity of the parameter space is not guaranteed, local optimality does not guarantee global optimality. Hence we run local optmizations from starting points on a grid with mesh side 0.2 over the subspace $[0,5]^4$ of the parameter space, and select our global maximum candidate as the local maximum with the highest entropy rate per unit time~~$\theta(M^*) \approx 1.92\,\mathrm{nats\,yr}^{-1}$. Unfortunately, this local maximum solution it is not guaranteed to be~~. Even though we cannot rigorously prove that our global maximum candidate $M_{\max} = M(\mathbf{u}, \mathrm{B}_{\max})$ as represented by the red dot in Fig. 6 with

$$\mathrm{B}_{\max} \approx \begin{pmatrix} -2.723 & 1.821 \\ 1.098 & -2.277 \end{pmatrix} \mathrm{yr}^{-1} \tag{67}$$

and $\theta_{\max} \approx 1.916$ is a global maximum~~entropy model in $\mathscr{M}$.~~

~~The nonidentifiability of the model from $\widehat{\Psi}$ alone is underlined by the fact that another system $\widetilde{M} = M(\mathbf{u}, \widetilde{\mathrm{B}}) \in \mathscr{M}$ with~~

$$\cancel{\widetilde{\mathrm{B}} = \begin{pmatrix} -2.00 & 2.00 \\ 1.00 & -3.00 \end{pmatrix} \mathrm{yr}^{-1}}$$

~~results in the same transfer function, but a different~~, we can clearly see that it is a good candidate. Increasing distance of local maximum parameters (panel (a)) and mean transit time (b) from the global maximum candidate lead to a decrease in entropy rate per unit time~~, that is, $\theta(\widetilde{M}) \approx 1.90\,\mathrm{nats\,yr}^{-1}$. 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. ??. 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~~ . Furthermore, local optimizations with starting points on the grid lead only to small improvements. A good choice of starting point on the grid is crucial to find a good global maximum candidate (c). Finally, the global maximum candidate for the entropy rate per unit time ~~of model $\widetilde{M}$~~ does not maximize the path entropy (d).

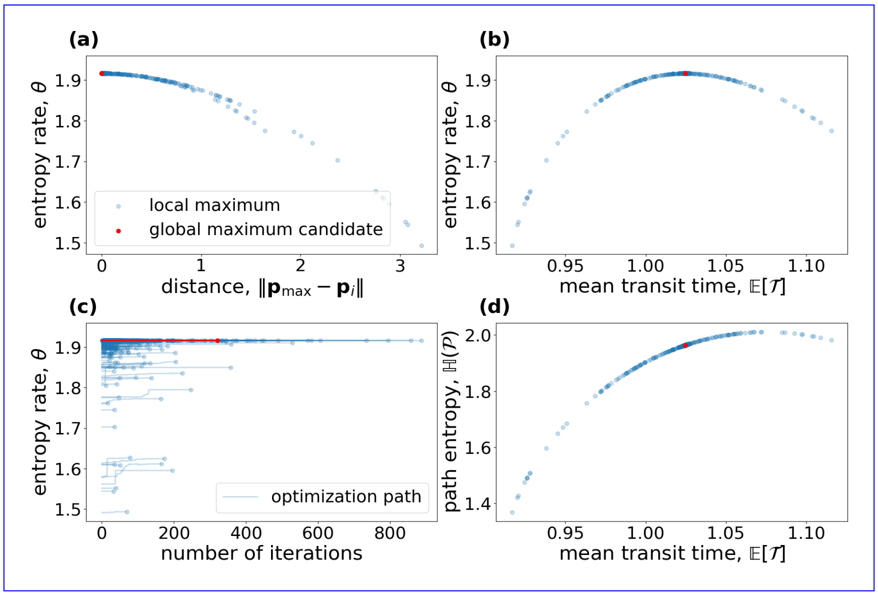


Fig. 6 ~~Evolution~~ Local maximizations of θ over a grid on a subspace of the ~~entropy~~ parameter space. For better visibility we chose randomly 1,000 grid points for the plot. Blue dots show local maxima found during the global maximization procedure starting on the grid. The red dot is associated to the global maximum candidate $M_{\max}$. (a) Entropy rate per unit time versus l_1-distance of ~~system~~ local maxima $\mathbf{p}_i$ parameters from the global maximum candidate parameters $\mathbf{p}_{\max}$. (b) Entropy rate per unit time versus mean transit time. (c) Paths of entropy rate per unit time during the ~~nonlinear optimization process for model identification~~ local maximizations on the grid. (d) Path entropy versus mean transit time

5 Discussion

Based on the stochastic path that a single particle takes through a deterministic 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 concept of path entropy as treated in the context of maximum caliber (MaxCal) ~~(Jaynes 1985)~~(Jaynes 1985 Roach 2020). 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. 4.1 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~~By virtue of its very mathematical definition (Eq. (1)), entropy is maximized when the system~~is highly symmetric~~'s symmetry is maximized. 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~~ . The Poisson process with intensity rate 1 is the one with maximum entropy rate, which follows directly from properties of the function $f(x) = x \log x$. 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, $\cancel{\mathbb{H}(M_2) \le \mathbb{H}(M_1)}$$\mathbb{H}(\mathscr{P}_2^*) \le \mathbb{H}(\mathscr{P}_1^*)$.

When we compute entropy values for actual carbon-cycle models (Sects. 4.2 and 4.3), we note that environmental or ~~biochemical factors impact the~~ eco-physiological factors might impact model entropies. For example, higher global surface temperatures might induce a higher global carbon cycle system speed $(1 < \xi << 6)$. This higher system speed reduces the uncertainty of the long-term future of entire paths of carbon atoms entering the terrestrial biosphere from the atmosphere. At the same time, it increases the entropy rate per unit time, that is, the uncertainty of the short-term future of carbon atoms already in the terrestrial biosphere.

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 two carbon-cycle models (Sects. 4.2 and 4.3) are well understood in equilibrium, hence they can serve as a means to better understand properties of the newly introduced entropy metrics. Once we understand entropy properties in dependence on general system properties, we can extrapolate this understanding to far more complex systems and make qualitative statements about their predictability without going into all model details. One major insight from those two examples is that, in general, slow heterogeneous systems are much harder to predict than fast homogeneous systems. Slowness increases the uncertainty of the duration of particle's stay in the system. Heterogeneity increases the uncertainty of a particle's sequence of visited compartments.

These simple insights allow us to understand modeling issues on grander scale, like the huge difference in the diversity of modeling approaches for carbon uptake by photosynthesis and the carbon cycle in soils. Both photosynthesis (García-Rodríguez et al. 2022) and soil carbon turnover (Manzoni and Porporato 2009) are modeled by many different approaches. However, in ecosystem models photosynthesis is almost exclusively represented based (Zaehle et al. 2014) on the Farquhar model (Farquhar et al. 1980), while soil carbon dynamics are represented by a great variety of models with very different structures (Friedlingstein et al. 2006). The latter leads to large variations in the prediction of future land carbon uptake (Friedlingstein et al. 2006; 2014). A comparison of carbon simulations from eleven model centers showed that across models global soil carbon varied more than twice as much as global net primary productivity (Todd-Brown et al. 2013) . Leaves have evolved to serve a specific purpose: to take up carbon from the atmosphere. Soils, on the other hand, have not been built to serve a specific purpose. They have evolved as a dumpster for material of which every soil agent tries to take advantage. This increases the heterogeneity of biogeochemical processes taking places in the soil. Furthermore, soil carbon turnover happens on much larger time scales than photosynthesis. While the photosynthetic apparatus operates on time scales in the order of split seconds to minutes, soil carbon cycling rates are in the order of decades to millenia. Consequently, the higher uncertainty of soil carbon cycling compared to photosynthetic carbon uptake is an inherent property of the system. Simply by the soil's heterogeneous and slow-cycling nature, the system has high inherent uncertainty, which hints at a theoretical limit that cannot be overcome by any model.

The example of model identification by MaxEnt in Sect. 4.4 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~~Owing to the nonlinear restrictions on the parameters in Eq. (64), the parameter space is probably not convex. Hence, local maxima are not guaranteed to be also globally optimal. The small size system size allows us to identify a reasonable global maximum candidate model by brute force, starting local maximizations on a grid over a parameter sub-space. Practical examples might include higher-dimensional systems and thus not be feasible for brute-force approaches. More sophisticated optimizations methods suitable for the particular problem at hand should then be applied.

6 Conclusions

~~Information content and complexity~~ The information content of autonomous compartmental systems in equilibrium can be assessed by the entropy of the path of particles traveling through ~~a~~ the 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~~possible descriptions, 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 us 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~~system uncertainty/predictability. 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 so far applied only to the inhomogeneous embedded jump chain as done by Ge et al. (2012) to incorporate also sojourn times in different compartments.

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.

Acknowledgements Funding was provided by the Max Planck Society and the German Research Foundation through its Emmy Noether Program (SI 1953/2–1) and the Swedish Research Council for Sustainable Development FORMAS, under grant 2018-01820.

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

Aoki I (1988) Entropy laws in ecological networks at steady state. Ecological Modelling 42(3):289–303, ISSN 0304-3800

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

Burnham K P, Anderson D R (2002) Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach. New York, Berlin, Heidelberg: Springer-Verlag

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

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

Fan J, Meng J, Ludescher J, Chen X, Ashkenazy Y, Kurths J, Havlin S, Schellnhuber H J (2021) Statistical physics approaches to the complex earth system. Physics Reports 896:1–84, ISSN 0370-1573, statistical physics approaches to the complex Earth system

Farquhar G D, von Caemmerer S v, Berry J A (1980) A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. planta 149:78–90

Friedlingstein P, Cox P, Betts R, Bopp L, von Bloh W, Brovkin V, Cadule P, Doney S, Eby M, Fung I, et al. (2006) Climate–carbon cycle feedback analysis: results from the c4mip model intercomparison. Journal of climate 19(14):3337–3353

Friedlingstein P, Meinshausen M, Arora V K, Jones C D, Anav A, Liddicoat S K, Knutti R (2014) Uncertainties in cmip5 climate projections due to carbon cycle feedbacks. Journal of Climate 27(2):511–526

García-Rodríguez L d C, Prado-Olivarez J, Guzmán-Cruz R, Rodríguez-Licea M A, Barranco-Gutiérrez A I, Perez-Pinal F J, Espinosa-Calderon A (2022) Mathematical modeling to estimate photosynthesis: A state of the art. Applied Sciences 12(11):5537

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

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

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

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

Golan A, Harte J (2022) Information theory: A foundation for complexity science. Proceedings of the National Academy of Sciences 119(33):e2119089119

Haddad W M (2013) A unification between dynamical system theory and thermodynamics involving an energy, mass, and entropy state space formalism. Entropy 15(5):1821–1846, ISSN 1099-4300

Haddad W M (2019) A Dynamical Systems Theory of Thermodynamics. Princeton: Princeton University Press

Haddad W M, Chellaboina V, Hui Q (2010) Nonnegative and Compartmental Dynamical Systems. Princeton University Press

Höge M, Wöhling T, Nowak W (2018) A primer for model selection: The decisive role of model complexity. Water Resources Research 54(3):1688–1715

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

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

Jaynes E T (1957b) Information Theory and Statistical Mechanics. ii. Physical Review 108(2):171–190

Jaynes E T (1985) Macroscopic prediction. In Complex Systems – Operational Approaches in Neurobiology, Physics, and Computers, Springer, 254–269

Jost J (2005) Dynamical systems: examples of complex behaviour. Berlin; New York: Springer

Manzoni S, Porporato A (2009) Soil carbon and nitrogen mineralization: Theory and models across scales. Soil Biology and Biochemistry 41(7):1355–1379

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

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

Morzy M, Kajdanowicz T, Kazienko P (2017) On measuring the complexity of networks: Kolmogorov complexity versus entropy. Complexity 2017:3250301

Neuts M F (1981) Matrix-geometric solutions in stochastic models: An algorithmic 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

Roach T N F (2020) Use and abuse of entropy in biology: A case for caliber. Entropy 22(12), ISSN 1099-4300

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 (2017) The muddle of ages, turnover, transit, and residence times in the carbon cycle. Global Change Biology 23(5):1763–1773, ISSN 1365-2486

Sierra C A, Quetin G R, Metzler H, Müller M (2023) A decrease in the age of respired carbon from the terrestrial biosphere and increase in the asymmetry of its distribution. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences in press

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

Todd-Brown K E, Randerson J T, Post W M, Hoffman F M, Tarnocai C, Schuur E A, Allison S D (2013) Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. Biogeosciences 10(3):1717–1736

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

Zaehle S, Medlyn B E, De Kauwe M G, Walker A P, Dietze M C, Hickler T, Luo Y, Wang Y P, El-Masri B, Thornton P, et al. (2014) Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate free-air CO_2 Enrichment studies. New Phytologist 202(3):803–822

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. (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~~

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

~~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~~

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

~~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~~

$$\mathbb{H}(Y_d) = -p\,\log p - (1-p)\,\log(1-p).$$

~~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~~

$$\mathbb{H}(Y_c) = 1 - \log\lambda$$

~~decreases with increasing λ.~~

~~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~~

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

~~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.~~

~~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 $p(y_1 \mid y_2)$ the conditional probability $\mathbb{P}(Y_1 = y_1 \mid Y_2 = y_2)$.~~

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

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

~~Note that $\mathbb{H}(Y_1 \mid Y_2) \leq \mathbb{H}(Y_1)$ 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~~

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

~~Let Y_3 be a third discrete random variable. Then~~

$$\mathbb{H}(Y_1, Y_2 \mid Y_3) = \mathbb{H}(Y_1 \mid Y_3) + \mathbb{H}(Y_2 \mid Y_1, Y_3).$$

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

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

~~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.~~

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

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

~~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) .~~

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

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

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

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

~~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.~~

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

$$\mathbb{H}_T(Z) = -\int f_T(z)\,\log f_T(z)\,\mathrm{d}\mu_T(z),$$

~~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.~~

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

$$\theta(Z) = \lim_{T\to\infty} \frac{1}{T}\,\mathbb{H}_T(Z)$$

~~if the limit exists.~~

A 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

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

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

Proposition A.1 Consider the set $\mathscr{M}_1$ of compartmental systems in equilibrium given by Eq. (6) 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

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

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

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

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

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

We rewrite the left hand side as

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

which leads to the constraints

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

The Lagrangian is now given by

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

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

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

and

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

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

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

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

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

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

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

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. (A.6) over $i \in S$,

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

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

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

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

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

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)

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

to rewrite the second term as

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

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 $j \in S$.

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

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

by Eq. (32). We now disturb B_{kl} for fixed $k,l \in S$ with $k \neq l$ by a sufficiently tiny ε, which may be 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 $z_j(\varepsilon) = z_j$, implying that the system's mean transit time remains unchanged, that is, $\mathbb{E}\left[\mathscr{T}(\varepsilon)\right] = \mathbb{E}\left[\mathscr{T}\right]$. The ε-disturbed H_3 is given by

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

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

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

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

Proposition A.2 Consider the set $\mathscr{M}_2$ of compartmental systems in equilibrium given by Eq. (6) with a predefined nonzero input vector $\mathbf{u}$ and a predefined positive steady-state vector $\mathbf{x}^*$. The compartmental system $M_2^* = M(\mathbf{u}, \mathrm{B}^*)$ with $\mathrm{B}^* = (B_{ij})_{i,j\in S}$ given by

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

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

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

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

An interchange of the indices and summing the two equations gives

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

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

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

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

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

and

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