

**1 Information content and maximum entropy of
2 compartmental systems in equilibrium**

3 Holger Metzler · Carlos A. Sierra

4
5 Received: Accepted:

6 Abstract Compartmental models are commonly used in different areas of science, particularly in modeling the cycles of carbon and other biogeochemical elements. The representation of these models as compartmental systems and assuming them to be in equilibrium is useful for comparisons of different model structures and parameterizations on a macroscopic scale. The interpretation of such models as continuous-time Markov chains allows a deeper model analysis on a microscopic scale. In particular we can assess 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 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 the maximum entropy principle (MaxEnt).

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

24 Mathematics Subject Classification (2000) 34A30 · 60J28 · 60K20 ·
25 92B05

Holger Metzler
Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Ulls väg 16, 756 51 Uppsala, Sweden, E-mail: holger.metzler@slu.de

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

26 1 Introduction

27 In a large variety of scientific fields such as systems biology, toxicology, pharma-
 28 cokinetics (Anderson, 1983), ecology (Eriksson, 1971; Rodhe and Björkström,
 29 1979; Matis et al, 1979; Manzoni and Porporato, 2009), hydrology (Nash, 1957;
 30 Botter et al, 2011; Harman and Kim, 2014), biogeochemistry (Manzoni and
 31 Porporato, 2009; Sierra and Müller, 2015), or epidemiology (Jacquez and Si-
 32 mon, 1993), models are based on the principle of mass conservation. In many
 33 cases such models are nonnegative dynamical systems that can be described by
 34 first-order systems of ordinary differential equations (ODEs) with strong struc-
 35 tural constraints. Such systems are called compartmental systems (Anderson,
 36 1983; Walter and Contreras, 1999; Haddad et al, 2010). We can classify such
 37 systems as combinations of linear/nonlinear and autonomous/nonautonomous
 38 (time-independent/time-dependent). For the sake of simplicity, most classical
 39 examples model natural processes by linear autonomous compartmental sys-
 40 tems (e.g., tracer kinetics, carbon cycle, leaky fluid tanks), often even in equi-
 41 librium. On the one hand, the simple structure of such systems allows a good
 42 understanding of undergoing processes in the modeled system. On the other
 43 hand, natural systems usually show highly complex interactions and depend
 44 on a constantly changing environment. Consequently, most of the time non-
 45 linear nonautonomous compartmental models (Kloeden and Pötzsche, 2013)
 46 are more appropriate to model natural systems.

47 Age and transit times are diagnostic tools of compartmental systems and
 48 have been widely studied for systems in and out of equilibrium (Eriksson, 1971;
 49 Bolin and Rodhe, 1973; Rasmussen et al, 2016; Sierra et al, 2016; Metzler and
 50 Sierra, 2018; Metzler et al, 2018). They help compare behavior and quality of
 51 different models. Nevertheless, structurally very different models might show
 52 very similar ages and transit times and might equally well represent given
 53 measurement data. If we are in the position to choose among such models,
 54 which is the one to select? This equifinality problem can be resolved by the
 55 maximum entropy principle (MaxEnt) (Jaynes, 1957a,b), a generic procedure
 56 to draw unbiased inferences from measurement or stochastic data (Pressé et al,
 57 2013). In order to apply MaxEnt to compartmental systems, some appropriate
 58 notion of entropy is required to measure the system's uncertainty or informa-
 59 tion content. Two classical examples in dynamical systems theory are the
 60 topological entropy and the Kolmogorov-Sinai/metric entropy. However, open
 61 compartmental systems are dissipative and by Pesin's theorem (Pesin, 1977)
 62 both metric and topological entropy vanish and cannot serve as a measure
 63 of uncertainty here. Alternatively, we can interpret compartmental systems
 64 as weighted directed graphs. Dehmer and Mowshowitz (2011) provide a com-
 65 prehensive overview of the history of graph entropy measures. Unfortunately,
 66 most of such entropy measures are based on the number of vertices, vertex
 67 degree, edges, or degree sequence (Trucco, 1956). Thus, they concentrate on
 68 only the structural information of the graph. There are also graph theoretic
 69 cal measures that take edges and weights into account by using probability
 70 schemes. Their drawback is that the underlying meaning of uncertainty be-

71 comes difficult to interpret because the assigned probabilities seem somewhat
72 arbitrary (Bonchev and Buck, 2005).

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

82 The manuscript is organized as follows. First we introduce basic notions
83 of information entropy and compartmental systems in equilibrium together
84 with their associated absorbing continuous-time Markov chain describing the
85 random path of one single particle through the system. Based on this Markov
86 chain we then we define three entropy quantities for compartmental systems
87 in equilibrium and adopt the MaxEnt theory. Afterwards we present the in-
88 troduced theory by means of simple generic examples and two carbon-cycle
89 models depending on changing environmental and biochemical parameters,
90 before we apply MaxEnt to a model identification problem.

91 2 Materials and methods

92 First, we introduce some basic notations and well-known properties of Shannon
93 information entropy of random variables and stochastic processes. Then we
94 present compartmental systems as a means to model material-cycle systems
95 that obey the law of mass balance. Then we consider such systems from a
96 single-particle point of view and define the path of a single particle through the
97 system along with its visited compartments, sojourn times, occupation times,
98 and transit time. Based on these basic structures of a path, we compute three
99 different types of entropy. For a better understanding, we provide a summary
100 of the desirable relations among the three different types:

- 101 (1) As a particle travels through the system, it jumps a certain number of
102 times to the next compartment until it finally jumps out of the system.
103 Between two jumps, the particle resides in some compartment. The *path*
104 *entropy* measures the entire uncertainty about the particles travel through
105 the system, including both the sequence of visited compartments and the
106 respective times spent there.
- 107 (2) The entire travel of the particle takes a certain time. In each unit time
108 interval before the particle leaves, uncertainties exist whether the particle
109 jumps, where it jumps, and even how often it jumps. The mean of these
110 uncertainties over the mean length of the travel interval is measured by
111 the *entropy rate per unit time*.
- 112 (3) Each jump comes with the uncertainties about which compartment will
113 be next and how long will the particle stay there. The *entropy rate per*

¹¹⁴ *jump* measures the average of these uncertainties with respect to the mean
¹¹⁵ number of jumps.

¹¹⁶ 2.1 Short summary of Shannon information entropy

¹¹⁷ We provide a short introduction of basic concepts of information entropy. For
¹¹⁸ a more detailed introduction along the lines of Cover and Thomas (2006) see
¹¹⁹ Section A

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

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

¹²⁰ is called the *Shannon information entropy (differential entropy)* of Y . Most
¹²¹ of the time we just say *entropy* and the precise meaning can be derived from
¹²² the context. The entropy's unit depends on the logarithmic base. For base 2
¹²³ the unit is bits and for the natural logarithm with base e the unit is nats.
¹²⁴ Throughout this manuscript we use the latter if not stated otherwise.

¹²⁵ The entropy $\mathbb{H}(Y)$ of a random variable Y has two intertwined interpretations. On the one hand, it is a measure of uncertainty, i.e., a measure of
¹²⁶ how difficult it is to predict the outcome of a realization of Y . On the other
¹²⁷ hand, $\mathbb{H}(Y)$ is also a measure of the information content of Y , i.e., a measure
¹²⁸ of how much information we gain once we learn about the outcome of a realization
¹²⁹ of Y . It is important to note that, even though their definitions and
¹³⁰ information theoretical interpretations are quite similar, the Shannon- and the
¹³¹ differential entropy have one main difference. The Shannon entropy is always
¹³² nonnegative, whereas the differential entropy can have negative values. While
¹³³ the Shannon entropy is an absolute measure of information and makes sense in
¹³⁴ its own right, the differential entropy is not an absolute information measure,
¹³⁵ is not scale-invariant, and makes sense only in comparison with the differential
¹³⁶ entropy of another random variable.

Panel A of Fig. 1 depicts the Shannon entropy of a Bernoulli random variable Y with $\mathbb{P}(Y = 1) = 1 - \mathbb{P}(Y = 0) = p$ with $p \in [0, 1]$ representing a coin toss with probability of heads equal to p . The closer p is to 1/2 the more difficult is it 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

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

is maximized. Panel B of Fig. 1 shows the differential entropy of an exponentially distributed random variable $Y \sim \text{Exp}(\lambda)$ 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 likelier

is an early exit of the particle, and the easier is it to predict the moment of exit. Hence, the differential entropy

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

- ¹³⁸ decreases with increasing λ .

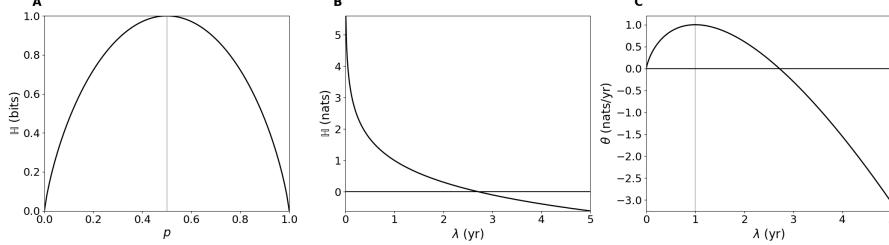


Fig. 1: A) Shannon entropy (logarithmic base 2) of a Bernoulli random variable depending on its success probability p . B) Differential entropy with logarithmic base e of an exponentially distributed random variable depending on its rate parameter λ . C) Entropy rate of a Poisson process with intensity rate λ .

Let Y_1 and Y_2 be two random variables with joint probability mass function (probability density function) $p(Y_1, Y_2)$. The *joint entropy* of Y_1 and Y_2 is given by

$$\mathbb{H}(Y_1, Y_2) = -\mathbb{E} [\log p(Y_1, Y_2)].$$

It is symmetric and $\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_1) + \mathbb{H}(Y_2)$ if Y_1 and Y_2 are independent. Furthermore, denote by $p(Y_1 | Y_2)$ the conditional probability mass function (conditional probability density function) of Y_1 given Y_2 . Then the *conditional entropy* of Y_1 given Y_2 is defined by

$$\mathbb{H}(Y_1 | Y_2) = -\mathbb{E} [\log p(Y_1 | Y_2)].$$

- ¹³⁹ Note that $\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_1) + \mathbb{H}(Y_2 | Y_1)$ and consequently $\mathbb{H}(Y_1 | Y_2) \leq \mathbb{H}(Y_1)$
¹⁴⁰ with equality if Y_1 and Y_2 are independent.

According to Bad Dumitrescu (1988) and Girardin and Limnios (2003) we can extend the concept of entropy to continuous-time stochastic processes $Z = (Z_t)_{\geq 0}$. We first define the entropy of Z on a finite time interval $[0, T]$ by

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

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. Note that by this definition we interpret the entire stochastic process Z on the interval $[0, T]$ as a single random variable on the space

$$\{z = (z_t)_{t \in [0, T]} : z_t \in \mathbb{R}\}.$$

Then the *entropy rate* of Z is defined by

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

¹⁴¹ if the limit exists.

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, Section 3.3), i.e.,

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

¹⁴² This entropy rate increases with $\lambda \in [0, 1]$, reaches its maximum at 1, and
¹⁴³ then it decreases (Fig. 1, panel C). This behavior is independent of the unit
¹⁴⁴ of λ , because it is based on the differential entropy of the exponential distri-
¹⁴⁵ bution and hence not scale-invariant. Consequently, it is no absolute measure
¹⁴⁶ of information content, but only useful in comparison to the entropy rates of
¹⁴⁷ other stochastic processes.

¹⁴⁸ 2.2 Compartmental systems in equilibrium

¹⁴⁹ Mass-balanced flow of material into a system, within the system and out of the
¹⁵⁰ system that consists of several compartments can modeled by so-called com-
¹⁵¹ partmental systems (Anderson, 1983). Following Jacquez and Simon (1993),
¹⁵² a *compartment* is an amount of some material that is kinetically homoge-
¹⁵³ neous. Compartments are usually also called *pools* or *boxes*. By kinetically
¹⁵⁴ homogeneous we mean that the material of a compartment is at all times ho-
¹⁵⁵ mogeneous; any material entering the compartment is instantaneously mixed
¹⁵⁶ with the material already there. Hence compartments are always *well-mixed*.
¹⁵⁷ One way to describe compartmental systems is by the d -dimensional linear
¹⁵⁸ ODE system

$$\frac{d}{dt} \mathbf{x}(t) = \mathbf{B} \mathbf{x}(t) + \mathbf{u}, \quad t > 0, \quad (1)$$

¹⁵⁹ with some initial condition $\mathbf{x}(0) = \mathbf{x}^0 \in \mathbb{R}^d$. The nonnegative vector $\mathbf{x}(t)$
¹⁶⁰ describes the amount of material in the different compartments at time t , the
¹⁶¹ nonnegative vector \mathbf{u} is the vector of external inputs to the compartments,
¹⁶² and the compartmental matrix $\mathbf{B} \in \mathbb{R}^{d \times d}$ describes the flux rates between
¹⁶³ the compartments and out of the system. To ensure that the system is mass
¹⁶⁴ balanced, we require the matrix \mathbf{B} be compartmental, i.e.,

- ¹⁶⁵ (i) all off-diagonal entries are nonnegative;
- ¹⁶⁶ (ii) all diagonal entries are nonpositive;
- ¹⁶⁷ (iii) all column sums are nonpositive.

The 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. We require additionally that at least one column sum of \mathbf{B} is strictly negative. This guarantees that the compartmental system is open in the sense that all material that enters the system will also leave the system at some point in time. The open compartmental system (1) has a unique steady-state or equilibrium compartment vector $\mathbf{x}^* = -\mathbf{B}^{-1} \mathbf{u}$ to which $\mathbf{x}(t)$ converges as $t \rightarrow \infty$, independently of the initial vector \mathbf{x}^0 . In this manuscript, we are interested only in systems that have already reached the equilibrium such that $\mathbf{x}(t) = \mathbf{x}^*$ for all $t \geq 0$. Note that also nonlinear systems, in which $\mathbf{B}(\mathbf{x})$, $\mathbf{u}(\mathbf{x})$ or both can depend on the system content \mathbf{x} , might reach a steady state $\mathbf{x}^* = -[\mathbf{B}(\mathbf{x}^*)]^{-1} \mathbf{u}(\mathbf{x}^*)$, in which case $\mathbf{B} = \mathbf{B}(\mathbf{x}^*)$ and $\mathbf{u} = \mathbf{u}(\mathbf{x}^*)$ are constant. A compartmental system in equilibrium given by Eq. (1) is fully characterized by \mathbf{u} and \mathbf{B} , and we denote it by $M = M(\mathbf{u}, \mathbf{B})$.

2.3 The one-particle perspective

While Eq. (1) describes the movement of bulk material through the system, compartmental systems in equilibrium can also be described probabilistically by considering the random path of a single particle through the system (Metzler and Sierra, 2018). If $X_t \in \mathcal{S} := \{1, 2, \dots, d\}$ 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 $\tilde{\mathcal{S}} := \mathcal{S} \cup \{d+1\}$. Its initial initial distribution is given by $\tilde{\beta} = (\beta_1, \beta_2, \dots, \beta_d, 0)^T$, where $\beta := \mathbf{u}/\|\mathbf{u}\|$ and $\beta_j = \mathbb{P}(X_0 = j)$ is the probability of the single particle to enter the system through compartment j . The superscript T denotes the transpose of the vector/matrix and $\|\mathbf{u}\| = \sum_{i=1}^d |u_i|$ denotes the l_1 -norm of the vector \mathbf{u} . The state-transition matrix of X is given by

$$\mathbf{Q} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{z}^T & 0 \end{pmatrix}, \quad (2)$$

and thus

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

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

$$\mathbb{P}(X_t = i \mid X_s = j) = (e^{(t-s)\mathbf{Q}})_{ij}, \quad s \leq t, \quad i, j \in \tilde{\mathcal{S}},$$

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. (1) are equivalent, we can write

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

197 2.4 The path of a single particle

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

$$\mathcal{P}(X) := ((Y_1 = X_0, T_1), (Y_2, T_2), \dots, (Y_{N-1}, T_{N-1}), Y_N = d+1), \quad (3)$$

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

$$\mathcal{N} := \inf \{n \in \mathbb{N} : Y_n = d+1\}$$

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

$$P_{ij} := \mathbb{P}(Y_{n+1} = i \mid Y_n = j) = \begin{cases} 0, & i = j \text{ or } \lambda_j = 0, \\ Q_{ij}/\lambda_j, & \text{else.} \end{cases} \quad (4)$$

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

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

204 is the diagonal matrix with the diagonal entries of \mathbf{Q} and \mathbf{I} denotes the identity
205 matrix of appropriate dimension. We define the matrix $\mathbf{P}_B := (P_{ij})_{i,j \in \mathcal{S}}$, then
206 $\mathbf{M} := (\mathbf{I} - \mathbf{P}_B)^{-1}$ is the *fundamental matrix* of Y , with $\mathbf{I} \in \mathbb{R}^{d \times d}$ denoting the
207 identity matrix. The entry M_{ij} denotes the expected numbers of visits to com-
208 partment i given that the particle entered the system through compartment
209 j . Consequently, the expected number of visits to compartment $i \in \mathcal{S}$ is given
210 by

$$N_i = \sum_{j=1}^d M_{ij} \beta_j = (\mathbf{M} \boldsymbol{\beta})_i = [(\mathbf{I} - \mathbf{P}_B)^{-1} \boldsymbol{\beta}]_i = (\mathbf{D} \mathbf{B}^{-1} \boldsymbol{\beta})_i = \frac{\lambda_i x_i^*}{\|\mathbf{u}\|} \quad (5)$$

and the total expected number of jumps is given by

$$\mathbb{E}[\mathcal{N}] = \sum_{i=1}^d (\mathbf{M} \boldsymbol{\beta})_i + 1 = \sum_{i=1}^d N_i + 1,$$

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

The last jump, \mathcal{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

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

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

$$f_{\mathcal{T}}(t) = \mathbf{z}^T e^{t \mathbf{B}} \boldsymbol{\beta}, \quad t \geq 0.$$

It can be shown (Metzler and Sierra, 2018, Section 3.2) that the mean or expected value of \mathcal{T} equals the resident time (Sierra et al, 2016) of system (1) in equilibrium and is given by total stocks over total fluxes, i.e.,

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

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

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

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

$$\mathbb{E} [O_j] = \frac{x_j^*}{\|\mathbf{u}\|}, \quad (6)$$

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

216 2.5 Path entropy, entropy rate per unit time, entropy rate per jump

217 The path $\mathcal{P}(X)$ given by Eq. (3) can be interpreted in three different ways.
218 Each of these ways leads to a different interpretation of the path's entropy.
219 First, we can look at \mathcal{P} as the result of bookkeeping of the absorbing continuous-
220 time Markov chain X , where as a sequence of pairs on the occasion of a jump
221 we note down the old compartment of the traveling particle and the associated
222 time the particle spent in this compartment. Second, we can consider the path
223 as a discrete-time process. In each time step n , we choose randomly a new
224 compartment Y_{n+1} and an associated sojourn time T_{n+1} of the particle in this
225 compartment. Third, we can look at \mathcal{P} as a single random variable with values
226 in the space of all possible paths. Based on the latter interpretation we now
227 derive the path entropy.

228 We are interested in the uncertainty/information content of the path $\mathcal{P}(X)$
229 of a single particle. Along the lines of Albert (1962), we construct a space \wp

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

Let $p = ((x_1, t_1), (x_2, t_2), \dots, (x_n, t_n), d + 1) \in \wp$ for some $n \in \mathbb{N}$. For $i \neq j$, we 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 $\mathcal{P} = \mathcal{P}(X)$ with respect to σ is given by*

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

$$p = ((x_1, t_1), (x_2, t_2), \dots, (x_n, t_n), d + 1) \in \wp.$$

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

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

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

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

$$p = ((x_1, t_1), (x_2, t_2), \dots, (x_n, t_n), d+1) \in \wp.$$

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

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

251 We make the according substitutions and the proof is finished.

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

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

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\mathcal{P}) \\ &= - \sum_{i=1}^d \beta_i \log \beta_i \\ &\quad + \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j) \right]. \end{aligned} \tag{7}$$

Proof Let X have the finite path representation

$$\mathcal{P} = \mathcal{P}(X) = ((Y_1, T_1), (Y_2, T_2), \dots, (Y_n, T_n), d+1)$$

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

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

We compute the expectation and get

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

Obviously, $\mathbb{E}[R_j(\mathcal{P})] = \mathbb{E}[O_j] = x_j^*/\|\mathbf{u}\|$ is the mean occupation time of compartment $j \in S$ by X . Furthermore, for $i \in \tilde{S}$ and $j \in S$ such that $i \neq j$, by Eqs. (5) and (4),

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

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

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

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

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

$$\mathbb{H}(X) = \mathbb{H}(\boldsymbol{\beta}) + \sum_{j=1}^d \mathbb{E}[O_j] \left(\sum_{i=1, i \neq j}^d \theta(\text{Poi}(B_{ij})) + \theta(\text{Poi}(z_j)) \right) \quad (8)$$

261 and

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\boldsymbol{\beta}) \\ &+ \sum_{j=1}^d \mathbb{E}[N_j] \left(\mathbb{H}(\text{Exp}(\lambda_j)) + \mathbb{H}(P_{1,j}, P_{2,j}, \dots, P_{d,j}, P_{d+1,j}) \right). \end{aligned} \quad (9)$$

Proof By virtue of Eq. (8) we replace $x_j^*/\|\mathbf{u}\|$ by $\mathbb{E}[O_j]$ in Eq. (7) and take into account that the entropy rate of a Poisson process with intensity rate λ equals $\lambda(1 - \log \lambda)$ to prove Eq. (8). To prove Eq. (9) we use Eq. (5) to replace $x_j^*/\|\mathbf{u}\|$ in Eq. (7) by $\mathbb{E}[N_j]/\lambda_j$ and obtain

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

262 Here, $(1 - \log \lambda_j)$ is the entropy of an exponential random variable with rate
263 parameter λ_j . Using definition (4) of P_{ij} we replace B_{ij}/λ_j by P_{ij} for $i \in \mathcal{S}$
264 and z_j/λ_j by $P_{d+1,j}$ ad finish the proof.

We now define the *path entropy of the compartmental system in equilibrium* M , given by Eq. (1), as the path entropy of its associated continuous-time Markov chain X , i.e.

$$\mathbb{H}(M) := \mathbb{H}(X) = \mathbb{H}(\mathcal{P}(X)).$$

- 265 For a one-dimensional compartmental system M_λ in equilibrium with rate
266 $\lambda > 0$ and positive external input given by

$$\frac{d}{dt} x(t) = -\lambda x(t) + u, \quad t > 0, \quad (10)$$

the entropy of the initial distribution vanishes, and we obtain

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

- 267 which equals the differential entropy $1 - \log \lambda$ of the exponentially distributed
268 mean transit time $\mathcal{T}_\lambda \sim \text{Exp}(\lambda)$, reflecting that the only uncertainty of the
269 particle's path in a one-pool system is the time of the particle's exit. The
270 exponential distribution with rate parameter λ is the distribution of the inter-
271 arrival time of a Poisson process with intensity rate λ . Hence, we can interpret
272 $\mathbb{H}(M_\lambda) = \lambda^{-1} \lambda (1 - \lambda)$ as the instantaneous Poisson entropy rate $\lambda(1 - \lambda)$
273 multiplied with the expected duration $\mathbb{E}[\mathcal{T}] = \lambda^{-1}$ of the particle's stay in the
274 system.

- 275 Migrating to a d -dimensional system, we can interpret $\mathbb{H}(M)$ as the en-
276 tropy of a continuous-time process in the light of Eq. (8) and as the entropy
277 of a discrete-time process in the light of Eq. (9). In both interpretations rep-
278 resents the first term $\mathbb{H}(\beta) = \mathbb{H}(X_0) = \mathbb{H}(Y_1)$ the uncertainty of the first pool
279 through which the particle enters the system. In the continuous-time inter-
280 pretation, the uncertainty of the subsequent travel is the weighted average of
281 the superposition of d Poisson processes describing the instantaneous uncer-
282 tainty of possible jumps of the particle inside the system, $\theta(\text{Poi}(B_{ij}))$, and
283 out of the system, $\theta(\text{Poi}(z_j))$, where the weights are the expected occupation
284 times of the different compartments $j \in \mathcal{S}$. In the discrete-time interpretation,
285 the subsequent travel's uncertainty is the average of uncertainties associated
286 to each pool, weighted by the number of visits to the respective pools. The
287 uncertainty associated to each pool comprises the uncertainty of the length
288 of the stay in the pool, $\mathbb{H}(\text{Exp}(\lambda_j))$, and the uncertainty of where to jump
289 afterwards, $\mathbb{H}(\{P_{ij} : i \in \tilde{\mathcal{S}}, j \in \mathcal{S}, i \neq j\})$.

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

$$\theta(M) = \frac{\mathbb{H}(M)}{\mathbb{E}[\mathcal{T}]}$$

and the *entropy rate per jump* by

$$\theta_J(M) = \frac{\mathbb{H}(M)}{\mathbb{E}[\mathcal{N}]}.$$

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

295 2.6 The maximum entropy principle (MaxEnt)

296 MaxEnt arose in statistical mechanics as a variational principle to predict
 297 the equilibrium states of thermal systems and later was applied to matters
 298 of information and as a general procedure to draw inferences based on self-
 299 consistency requirements (Pressé et al, 2013). Its relationships to information
 300 theory and stochastics were established by Jaynes (1957a,b). The general idea
 301 is to identify the most uninformed probability distribution to represent some
 302 given data in the sense that the maximum entropy distribution, constrained to
 303 given data, uses the information provided by the data only and nothing else.
 304 This approach ensures that no additional subjective information creeps into
 305 the distribution. The goal of this section is to transfer MaxEnt to compart-
 306 mental systems in order to identify the compartmental system that represents
 307 our state of knowledge best in different situations, and at the same time get a
 308 better understanding of the introduced entropy measures.

Example 1 Consider the set \mathcal{M}_1 of equilibrium compartmental systems (1) with a predefined nonzero input vector \mathbf{u} , a predefined mean transit time $\mathbb{E}[\mathcal{T}]$, 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 $M_1^* \in \mathcal{M}_1$ such that the entropy rate per unit time $\theta(M_1^*)$, or equivalently the path entropy $\mathbb{H}(\mathcal{P}(M_1^*))$, is maximized. We can show (see Proposition B.1) that the compartmental system $M_1^* = M(\mathbf{u}, \mathbf{B})$ with

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

309 where $\lambda = d - 1 + 1/\mathbb{E}[\mathcal{T}]$, is the maximum entropy model in \mathcal{M}_1 . In the
 310 special case $d = 1$ for a one-dimensional compartmental system, we obtain
 311 $B = -1/\mathbb{E}[\mathcal{T}]$. Since in this case $\mathcal{T} \sim \text{Exp}(-B)$, we see that the exponential
 312 distribution is the maximum entropy distribution in the class of all nonnegative
 313 continuous probability distributions with fixed expected value. This special
 314 case is very well known (Cover and Thomas, 2006, Example 12.2.5).

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

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

315 is the maximum entropy model in \mathcal{M}_2 (see Proposition B.2).

316 2.7 Structural model identification via MaxEnt

317 Suppose we observe a natural system and conduct measurements from which
 318 we try to construct a linear autonomous compartmental model in equilibrium
 319 that represents the observed natural system as well as possible. The first ques-
 320 tion that arises is the one for the number of compartments the model should
 321 ideally have. MaxEnt cannot be helpful here because by adding more and more
 322 compartments we can theoretically increase the entropy of the model indefi-
 323 nitely. Consequently, the problem of finding the right dimension of system (1)
 324 has to be solved by other means. One way to do it is to analyze an impulse
 325 response function of the system and its Laplace transform, i.e. the transfer
 326 function of the system, and identify the most dominating frequencies. The im-
 327 pulse response or the transfer function might be possible to obtain by tracer
 328 experiments (Anderson, 1983; Walter, 1986).

329 In Anderson (1983, Chapter 16) the *structural identification problem* of
 330 linear autonomous systems is described as follows. Suppose we are interested
 331 in determining a d -dimensional system of form (1). We are interested in sending
 332 an impulse into the system at time $t = 0$ and analyzing its further behavior.
 333 To that end, we rewrite the system to

$$\begin{aligned} \frac{d}{dt} \mathbf{x}(t) &= \mathbf{B} \mathbf{x}(t) + \mathbf{A} \mathbf{u}, \quad t \geq 0, \\ \mathbf{x}(0) &= \mathbf{0} \in \mathbb{R}^d, \\ \mathbf{y}(t) &= \mathbf{C} \mathbf{x}(t), \quad t \geq 0. \end{aligned} \tag{11}$$

Note that the roles of \mathbf{A} and \mathbf{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* \mathbf{A} , which defines how the input vector enters the system. Then we decide which compartments we can observe to determine the *output connection matrix* \mathbf{C} . The experiment is now to inject an impulse into the system and to record the output function $\mathbf{y}(t) = \mathbf{C} \mathbf{x}(t)$. Bellman and Åström (1970) pointed out that the input-output relation is given by

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

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

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

or in the *transfer function*

$$\hat{\Psi}(s) := C(sI - B)^{-1} A, \quad s \geq 0,$$

which is the Laplace transform matrix of Ψ . Consequently, all identifiable parameters of A , B , and C must be identified through Ψ or $\hat{\Psi}$. Difficulties arise because the entries of the matrices Ψ and $\hat{\Psi}$ are usually nonlinear expressions of the elements of A , B , and C . We call system (11) *identifiable* if this nonlinear system of equations has a unique solution (A, B, C) for given Ψ or $\hat{\Psi}$. Otherwise the system is called *nonidentifiable*. Usually, the matrices A and C are already known from the experiment's setup. What remains is to identify the compartmental matrix B , and this can be done by MaxEnt.

3 Results/Examples

First, we apply the presented theory to some equilibrium compartmental models with very simple structure in order to get some grasp on the new entropy concept. Then we compute entropy quantities for two carbon-cycle models in dependence on environmental and biochemical parameters. Afterwards, we apply MaxEnt to solve an equifinality problem in model selection.

3.1 Simple examples

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

3.2 A linear autonomous global carbon-cycle model

We consider the global carbon-cycle model introduced by Emanuel et al (1981) (Fig. 2). The model comprises five compartments: non-woody tree parts x_1 , woody tree parts x_2 , ground vegetation x_3 , detritus/decomposers x_4 , and active soil carbon x_5 . We introduce an environmental rate modifier ξ which controls the speed of the system. This parameter could potentially increase

Structure	$\frac{d}{dt} \mathbf{x}(t)$	θ_J	$\mathbb{E}[N]$	θ	$\mathbb{E}[T]$	$\mathbb{H}(\mathcal{P})$
	$-\lambda x + 1$	$0.5(1 - \log \lambda)$	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
	$\begin{pmatrix} -1 & 0.5 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	1.08	5.00	1.35	<u>4.00</u>	<u>5.39</u>
	$\begin{pmatrix} -1 & 0.5 \\ 0.5 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	<u>1.36</u>	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	<u>2.10</u>	1.00	2.10

Table 1: Overview of different entropy measures of simple models with different structures. The columns from left to right represent a schematic of the model, its mathematical representation, its entropy rate per jump, its mean number of jumps, its entropy rate per unit time, its mean transit time, and its path entropy. Underlined numbers are the highest values per column.

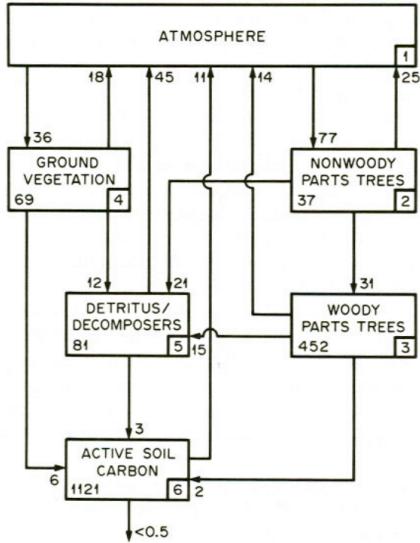


Fig. 2: Schematic of the linear autonomous global carbon cycle model in steady state introduced by Emanuel et al (1981). The model comprises five compartments: non-woody tree parts x_1 (2; 37 PgC), woody tree parts x_2 (3; 452 PgC), ground vegetation x_3 (4; 69 PgC), detritus/decomposers x_4 (5; 81 PgC), and active soil carbon x_5 (6; 1,121 PgC). The atmosphere (1) is considered to be outside of the modeled system but provides the system with external inputs and receives external outputs from it. Numbers next to arrows indicate fluxes between compartments in PgC yr^{-1} . (Figure extracted from Emanuel et al 1981)

and speed up the system with increasing global Earth surface temperature. For given ξ , the equilibrium model $M_\xi = M(\mathbf{u}, \mathbf{B}_\xi)$ is given by

$$\mathbf{u} = (77; 0; 36; 0; 0)^T \text{ PgC yr}^{-1}$$

and

$$\mathbf{B}_\xi = \xi \begin{pmatrix} -77/37 & 0 & 0 & 0 & 0 \\ 31/37 & -31/452 & 0 & 0 & 0 \\ 0 & 0 & -36/69 & 0 & 0 \\ 21/37 & 15/452 & 12/69 & -48/81 & 0 \\ 0 & 2/452 & 6/69 & 3/81 & -11/1121 \end{pmatrix} \text{ yr}^{-1},$$

where the numbers are chosen as in Thompson and Randerson (1999). The input vector is expressed in units of petagrams of carbon per year (PgCyr^{-1}) and the fractional transfer coefficients in units of per year (yr^{-1}). Because \mathbf{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 is 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. Nevertheless, the entropy rate per jump (panel F) decreases with increasing ξ , because the path entropy of the system decreases. The entropy rate per unit time increases until $\xi \approx 6$ while the mean transit time decreases faster than the path entropy, and then the trend turns around and the entropy rate per unit time decreases (panel E). Orange lines in panel D and E show the respective entropy values for a one-pool system $M_\lambda = M((77 + 36) \text{ PgCyr}^{-1}, \lambda)$ with the same mean transit time, i.e. $\lambda^{-1} = \mathbb{E}[\mathcal{T}_\xi]$. The blue and orange lines intersect at $\xi \approx 4.31$.

3.3 A nonlinear autonomous soil organic matter decomposition model

Consider the nonlinear two-compartment carbon-cycle model $M_\varepsilon = M(\mathbf{u}, \mathbf{B}_\varepsilon)$ described by Wang et al (2014) (Fig. 4) and given by

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

where $\mathbf{x}(t) = (C_s, C_b)^T(t)$. We denote by C_s and C_b soil organic carbon and soil microbial biomass (gC m^{-2}), respectively, by ε the carbon use efficiency or fraction of assimilated carbon that is converted into microbial biomass (unitless), by μ_b the turnover rate of microbial biomass per year (yr^{-1}), by F_{NPP} the carbon influx into soil ($\text{gC m}^{-2} \text{ yr}^{-1}$), 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 (gC m^{-2}), respectively.

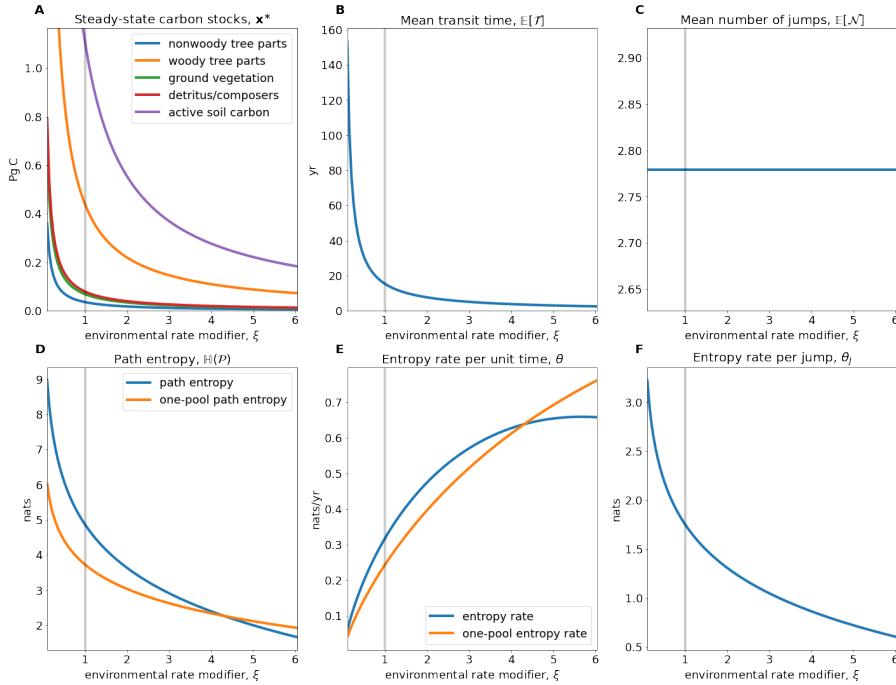


Fig. 3: A) Equilibrium carbon stocks. B)–F) Entropy related quantities of the global carbon cycle model introduced by Emanuel et al (1981) in dependence on the environmental rate coefficient ξ (blue lines). Orange lines correspond to the quantities derived from a one-pool model with the same speed. Vertical gray lines show $\xi = 1$, the original speed of the model.

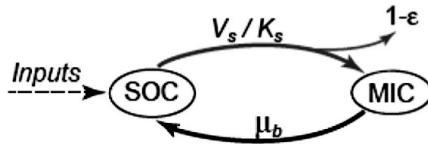


Fig. 4: Scheme of the nonlinear autonomous carbon cycle model introduced by Wang et al (2014). The two compartments C_s and C_b are here denoted by SOC (substrate organic carbon) and MIC (microbial biomass carbon), the external input flux F_{NPP} is denoted by *Inputs*, the maximum rate of soil carbon assimilation by V_s , the half saturation constant by K_s , the carbon use efficiency by ε , and the turnover rate of microbial biomass by μ_b , respectively. (Figure extracted from Wang et al (2014))

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

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

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

$$\lambda(\mathbf{x}) = \frac{C_b V_s}{C_s + K_s}. \quad (12)$$

386 From Wang et al (2014) we take the parameter values $F_{\text{NPP}} = 345.00 \text{ gC m}^{-2} \text{ yr}^{-1}$,
387 $\mu_b = 4.38 \text{ yr}^{-1}$, and $K_s = 53,954.83 \text{ gC m}^{-2}$. Since the description of V_s is
388 missing in the original publication, we let it be equal to 59.13 yr^{-1} to ap-
389 proximately meet the given steady-state contents $C_s^* = 12,650.00 \text{ gC m}^{-2}$ and
390 $C_b^* = 50.36 \text{ gC m}^{-2}$ for the original value $\varepsilon = 0.39$. Otherwise we leave the
391 carbon use efficiency ε as a free parameter.

392 In contrast to the system from the first example, this system exhibits a
393 feedback. This feedback results from dead soil microbial biomass being con-
394 sidered as new soil organic matter. The feedback can also be recognized by
395 noting that \mathbf{B} is not triangular. For every value of ε the system has a different
396 steady state (Fig. 5, panel A). The higher the value of ε , the lower the equilib-
397 rium substrate organic carbon and the higher the microbial biomass carbon.
398 Caused by the model's nonlinearity expressed in Eq. (12), the system speed
399 increases and the mean transit time goes down (panel B). At the same time,
400 higher carbon use efficiency increases the probability of the carbon atom to
401 be reused more often, hence the mean number of jumps increases (panel C),
402 making the entropy rate per jump decrease (panel F). Even though the average
403 paths become shorter, with increasing carbon use efficiency the path entropy
404 increases as well for most values of ε . This has two reasons. First, the uncer-
405 tainty of where to jump from C_s increases, this uncertainty decreases then
406 for $\varepsilon > 0.5$. Second, the rate $-B_{11}$ of leaving the substrate pool is increasing
407 and smaller than 1. The corresponding Poisson process reaches its maximum
408 entropy rate at an intensity rate equal to 1 (Fig. 1, panel C), here at $\varepsilon \approx 0.926$.
409 This is also reflected in the entropy rate per unit time (panel D). The maxi-
410 mum does not exactly occur at $\varepsilon = 0.926$, because the times that the particle
411 stays in the different pools also depends on ε . For $\varepsilon > 0.926$ both the path
412 entropy and the entropy rate rapidly decline as both the jump uncertainty
413 and the Poisson entropy rate decline sharply. Considering a one-pool system
414 $M_\lambda = M(345.00 \text{ gC m}^{-2} \text{ yr}^{-1}, 1/\mathbb{E}[\mathcal{T}_\varepsilon])$ with the same mean transit time, we
415 recognize only small sensitivity of the path entropy on ε , because the contrary
416 effects on $\mathbb{E}[\mathcal{T}]$ and θ mostly balance out (orange lines in panels D and E).

417 3.4 Model identification via Maxent

418 The following example is inspired by Anderson (1983, Example 16 C). It shows
419 how MaxEnt can help take a decision which model to use if not all parameters

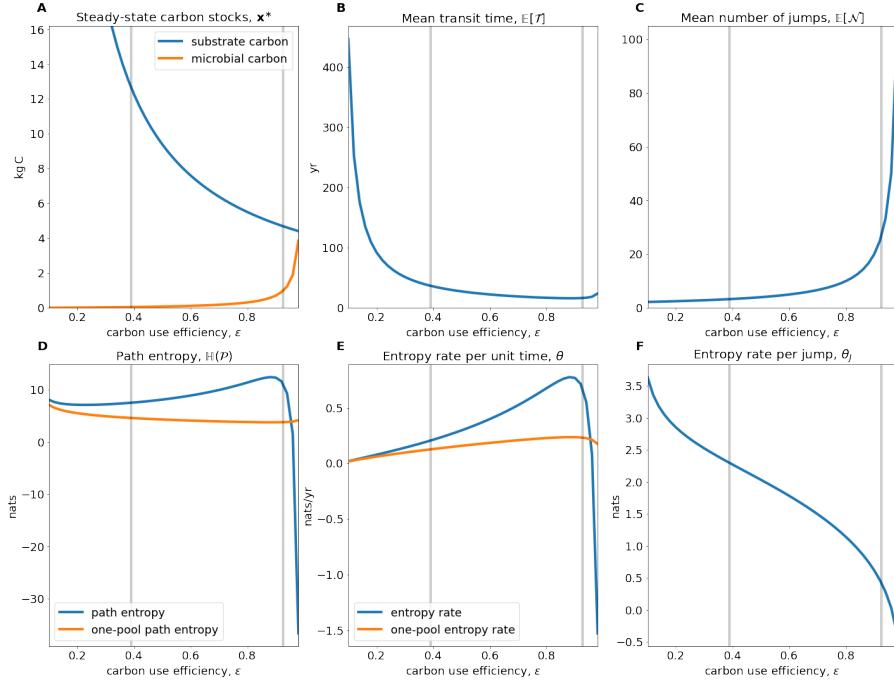


Fig. 5: A) Equilibrium carbon stocks. B)-F) Entropy related quantities of the global carbon cycle model introduced by Wang et al (2014) in dependence on the microbial carbon use efficiency ε (blue lines). Orange lines correspond to the quantities derived from a one-pool model with the same speed. The left vertical gray lines show $\varepsilon = 0.39$, the original carbon use efficiency of the model, the right $\varepsilon = 0.926$, the carbon use efficiency value with the maximum entropy rate of the Poisson process associated with C_s .

420 can be uniquely determined from the transfer function $\hat{\Psi}$. We are interested
 421 in determining the entries of the compartmental matrix B belonging to the
 422 2-dimensional equilibrium compartmental system

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (t) = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (t) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{gC yr}^{-1}, \quad t > 0. \quad (13)$$

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

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

⁴²³ where

$$\begin{aligned}\gamma_1 &= B_{12} + z_2, \\ \gamma_2 &= B_{21} + z_1 + B_{12} + z_2, \\ \gamma_3 &= z_1 B_{12} + z_1 z_2 + B_{21} z_2.\end{aligned}\tag{14}$$

⁴²⁴ We assume that $\hat{\Psi}$ is known from measurements, i.e., γ_1 , γ_2 , and γ_3 are known
⁴²⁵ impulse response parameters. We have the four unknown parameters B_{11} , B_{12} ,
⁴²⁶ B_{21} , and B_{22} , or equivalently, B_{12} , B_{21} , z_1 , and z_2 , but only three equations to
⁴²⁷ determine them. Consequently, the system is nonidentifiable and there remains
⁴²⁸ a class \mathcal{M} of models which all satisfy Eq. (14). Which model out of \mathcal{M} are we
⁴²⁹ 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 $M^* = (\mathbf{u}, \mathbf{B}^*)$ such that

$$M^* = \arg \max_{M \in \mathcal{M}} \theta(M).$$

⁴³⁰ We maximize the entropy rate per unit time here instead of the path entropy,
⁴³¹ because by slowing down the model, we could potentially increase its mean
⁴³² transit time and with it its path entropy indefinitely.

Let us turn to a numerical example in which we suppose to be given $\gamma_1 = 3 \text{ yr}^{-1}$, $\gamma_2 = 5 \text{ yr}^{-1}$, and $\gamma_3 = 4 \text{ yr}^{-1}$. A nonlinear optimization algorithm with the arbitrarily chosen initial values $B_{12} = 3 \text{ yr}^{-1}$, $B_{21} = 0 \text{ yr}^{-1}$, $z_1 = 1 \text{ yr}^{-1}$, and $z_2 = 1 \text{ yr}^{-1}$ ends approximately with the terminal compartmental matrix

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

⁴³³ and the terminal entropy rate per unit time $\theta(M^*) \approx 1.92 \text{ nats yr}^{-1}$. Unfortunately, this local maximum solution it is not guaranteed to be a global
⁴³⁴ maximum entropy model in \mathcal{M} .

The nonidentifiability of the model from $\hat{\Psi}$ alone is underlined by the fact that another system $\tilde{M} = (\mathbf{u}, \tilde{\mathbf{B}}) \in \mathcal{M}$ with

$$\tilde{\mathbf{B}} = \begin{pmatrix} -2.00 & 2.00 \\ 1.00 & -3.00 \end{pmatrix} \text{ yr}^{-1}$$

⁴³⁶ results in the same transfer function, but a different entropy rate per unit
⁴³⁷ time, i.e., $\theta(\tilde{M}) \approx 1.90 \text{ nats yr}^{-1}$ (dashed line in Fig. 6).

⁴³⁸ 4 Discussion

⁴³⁹ Based on the path that a particle takes through a compartmental system, we
⁴⁴⁰ introduced three types of entropy based on Shannon information theory. The
⁴⁴¹ entropy of the particle's entire path through the system is the central concept,
⁴⁴² and the entropy rates per unit time and per jump are consistently derived from
⁴⁴³ it. Even though we call $\mathbb{H}(\mathcal{P})$ path entropy and identify models by maximizing

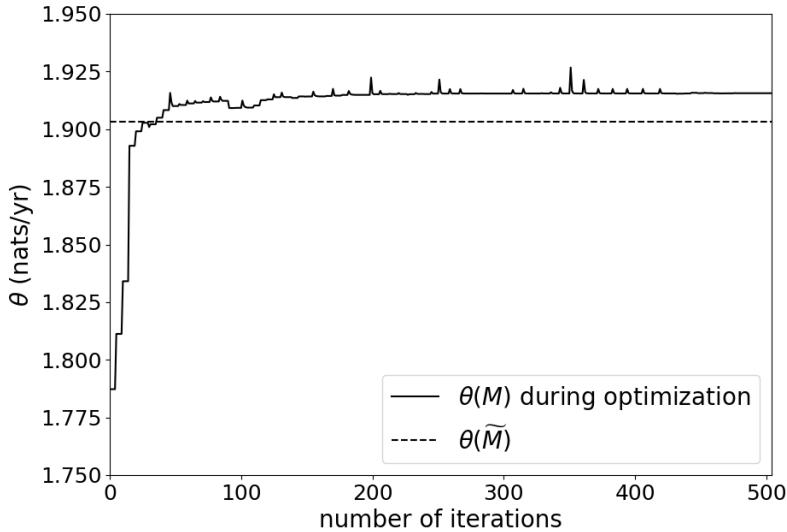


Fig. 6: Entropy rate per unit time of system (13). The solid curve shows the evolution of the entropy rate per unit time during the nonlinear optimization process. Peaks higher than the terminal value show attempts of the optimization algorithm that do not perfectly satisfy all constraints. The dashed line shows the entropy rate per unit time of model \tilde{M} .

444 it, it is different from the concept of path entropy as treated in the context of
 445 maximum caliber (MaxCal) (Jaynes, 1985). We maximize here the Shannon
 446 entropy of a single particle's microscopic path through a compartmental sys-
 447 tem by means of an absorbing continuous-time Markov chain whose transition
 448 probabilities are already determined by the macroscopic equilibrium state of
 449 the system. As discussed by Pressé et al (2013), MaxCal interprets the path
 450 entropy as a macroscopic system property to be maximized in order to iden-
 451 tify a time-dependent trajectory of the entire dynamical system, not just one
 452 single particle.

453 In the field of soil carbon cycle modeling, Ågren (2021) recently applied the
 454 maximum entropy principle to identify the quality distribution in the frame-
 455 work of continuous-quality theory. Given only the nonnegative mean quality,
 456 an application of MaxEnt leads to an exponential quality distribution, be-
 457 cause under these circumstances the exponential distribution is the maximum
 458 entropy distribution. The path entropy generalizes this approach to several
 459 interconnected compartments and jumps between them, while each sojourn
 460 time in a compartment is exponentially distributed.

461 From the simple examples in Section 3.1 we can see that models can be
 462 ordered differently in terms of uncertainty, depending on whether the interest

463 is in the uncertainty of the entire path or in some average uncertainty rate.
 464 For applications of MaxEnt without restrictions on the transit time, it is often
 465 useful to maximize an entropy rate because by slowing the system down more
 466 and more, the path entropy can potentially be increased indefinitely and there
 467 is no way to find a maximum path entropy model.

468 Usually, entropy is maximized when the system is highly symmetric. This
 469 is indicated by the Bernoulli entropy (Fig. 1, panel A) and supported by
 470 Example 1. Intuitively, this result is obvious. The system has a high sym-
 471 metry, the particle is equally likely to jump among different pools, and the
 472 Poisson process with intensity rate 1 is the one with maximum entropy rate.
 473 Furthermore, the resulting rates $z_j = 1/\mathbb{E}[\mathcal{T}]$ of leaving the system are chosen
 474 such that the mean transit time constraint is fulfilled. In Example 2, the sym-
 475 metry is broken by the additional restriction of a given steady-state vector.
 476 Consequently, $\mathbb{H}(M_2) \leq \mathbb{H}(M_1)$ with equality if the system content is equally
 477 distributed among all compartments.

478 When we compute entropy values for actual carbon-cycle models (Sec-
 479 tions 3.2 and 3.3), we note that environmental or biochemical factors impact
 480 the model entropies. Furthermore, in panels D and E of Fig. 3 we see that be-
 481 fore the break-even point of $\xi \approx 4.31$ the path of the Emanuel model is harder
 482 to predict than the path (i.e. the exit time of the particle) of a one-pool model
 483 with the same mean transit time. After this point of break even, the path of
 484 the Emanuel model with five compartments is easier to predict than only the
 485 transit time in a one-pool model. The reason is that as the system becomes
 486 faster, the differential entropy of the sojourn times in slow pools decreases so
 487 fast that at some point the sojourn times in slow pools visited by few par-
 488 ticles becomes rather unimportant. In consequence this means that after the
 489 break-even point, i.e. for a sufficiently fast system, a one-pool model is too
 490 biased on the slow-cycling paths, while fast paths are dominating the system.
 491 The path of a detailed model that separates fast from slow paths is then even
 492 easier to predict than a one-pool model path, even though the paths look more
 493 complicated.

494 The example of model identification by MaxEnt in Section 3.4 shows a ma-
 495 jor difference to the more artificial previous maximum entropy examples. The
 496 given constraints do not tell us enough about the structure of the model class
 497 \mathcal{M} to ensure that an identified local maximum is also a global maximum. It
 498 might be possible that with different initial values the optimization algorithm
 499 finds another local maximum model with higher entropy rate. This example is
 500 only supposed to give a first impression of how the maximum entropy principle
 501 can be used in combination with entropy rates or path entropy in similar sit-
 502 uations. Practical examples usually have a high level of complexity such that
 503 existence and uniqueness of a maximum entropy model have to be studied on
 504 a case-by-case basis.

505 5 Conclusions

506 We aimed at applying MaxEnt to models of mass-balanced dynamical systems,
507 so-called compartmental systems, in order to solve the problem of equifinality.
508 Since two of the most popular entropy measures for dynamical systems,
509 namely topological and metric entropy, vanish for such systems and cannot
510 serve as foundation for MaxEnt, we introduced another concept. We inter-
511 preted the system from a one-particle point of view and analyzed it in terms
512 of information entropy. When a particle moves through the system, it creates
513 a path from the time of its entry until the time of its exit. We can describe
514 this path in three ways: (1) as a random variable in the path space; (2) as a
515 continuous-time stochastic process representing the visited compartments; (3)
516 as a discrete sequence of pairs consisting of visited compartments and associ-
517 ated sojourn times. Based on these three ways, we introduced for systems in
518 equilibrium (1) the entropy of the entire path, (2) the entropy rate per unit
519 time, and (3) the entropy rate per jump. These three interpretations lead to
520 the same path entropy, which is a measure of how difficult the path of the
521 particle is to predict at the moment of entry. The concept of path entropy for
522 compartmental systems sets the foundation for several future research direc-
523 tions.

524 So far, the path entropy is developed only for systems in equilibrium. Since
525 most natural systems are far away from equilibrium, an extension of the path
526 entropy concept to nonautonomous compartmental systems is desirable. This
527 can be done by building on the concept of the entropy rate per unit time as
528 an instantaneous uncertainty and interpreting nonautonomous compartmental
529 systems as inhomogeneous Markov chains. This would allow an extension of
530 MaxCal applied only to the inhomogeneous embedded jump chain as done by
531 Ge et al (2012).

532 The path entropy might allow us in the future to assess theoretical limits
533 in the reduction of model uncertainty and to identify bottlenecks in modeling
534 theory. As we have seen, the path entropy is higher for slow systems. Conse-
535 quently, the detailed paths of particles through slow systems are more difficult
536 to predict than through fast systems. The concept of path entropy supports the
537 hypothesis that most uncertainty in land carbon uptake (Friedlingstein et al,
538 2006, 2014) is caused by the soil, because the soil contains a huge amount of
539 the global carbon and soil carbon turnover is comparatively slow.

540 We can also interpret the path entropy as a measure of the information
541 content of a compartmental system, because each particle's path through the
542 system produces some amount of information. At the same time measurement
543 data sets from natural systems contain a certain amount of information. There
544 is general lack of insight into the link between these two types of information. Is
545 a certain model capable of producing paths with sufficient information content
546 such that it is adequate to be used to reproduce available data? By introducing
547 the concept of path entropy to compartmental systems, we made a first crucial
548 step toward closing this knowledge gap.

549 6 Acknowledgements

550 Funding was provided by the Max Planck Society and the German Research
551 Foundation through its Emmy Noether Program (SI 1953/2-1) and the Swedish
552 Research Council for Sustainable Development FORMAS, under grant 2018-
553 01820.

554 References

- 555 Ågren GI (2021) Investigating soil carbon diversity by combining the maxi-
556 mum entropy principle with the q model. *Biogeochemistry* 153(1):85–94
- 557 Albert A (1962) Estimating the Infinitesimal Generator of a Continuous
558 Time, Finite State Markov Process. *The Annals of Mathematical Statis-
559 tics* 33(2):727–753
- 560 Anderson DH (1983) Compartmental modeling and tracer kinetics, vol 50.
561 Springer Science & Business Media
- 562 Bad Dumitrescu ME (1988) Some informational properties of markov pure-
563 jump processes. *Časopis pro pěstování matematiky* 113(4):429–434
- 564 Bellman R, Åström KJ (1970) On structural identifiability. *Mathematical Bio-
565 sciences* 7(3-4):329–339
- 566 Bolin B, Rodhe H (1973) A note on the concepts of age distribution and transit
567 time in natural reservoirs. *Tellus* 25(1):58–62
- 568 Bonchev D, Buck GA (2005) Quantitative Measures of Network Complexity.
569 In: Complexity in Chemistry, Biology, and Ecology, Springer, pp 191–235
- 570 Botter G, Bertuzzo E, Rinaldo A (2011) Catchment residence and travel time
571 distributions: The master equation. *Geophysical Research Letters* 38(11)
- 572 Cover TM, Thomas JA (2006) Elements of Information Theory, 2nd edn. Wiley
- 573 Dehmer M, Mowshowitz A (2011) A history of graph entropy measures. *Infor-
574 mation Sciences* 181(1):57–78
- 575 Doob JL (1953) Stochastic Processes, vol 7. Wiley, New York
- 576 Emanuel WR, Killough GG, Olson JS (1981) Modelling the Circulation of
577 Carbon in the World's Terrestrial Ecosystems. In: Carbon Cycle Modelling,
578 SCOPE 16, John Wiley and Sons, pp 335–353
- 579 Eriksson E (1971) Compartment Models and Reservoir Theory. *Annual Review
580 of Ecology and Systematics* 2:67–84
- 581 Friedlingstein P, Cox P, Betts R, Bopp L, Von Bloh W, Brovkin V, Cadule
582 P, Doney S, Eby M, Fung I, Bala G, John J, Jones C, Joos F, Kato T,
583 Kawamiya M, Knorr W, Lindsay K, Matthews HD, Raddatz T, Rayner P,
584 Reick C, Roeckner E, Schnitzler KG, Schnur R, Strassmann K, Weaver AJ,
585 Yoshikawa C, Zeng N (2006) Climate-carbon cycle feedback analysis: Results
586 from the C4MIP model intercomparison. *Journal of Climate* 19(14):3337–
587 3353
- 588 Friedlingstein P, Meinshausen M, Arora VK, Jones CD, Anav A, Liddicoat
589 SK, Knutti R (2014) Uncertainties in CMIP5 Climate Projections due to
590 Carbon Cycle Feedbacks. *Journal of Climate* 27(2):511–526

- 591 Gaspard P, Wang XJ (1993) Noise, chaos, and (ε, τ) -entropy per unit time.
592 Physics Reports 235(6):291–343
- 593 Ge H, Pressé S, Ghosh K, Dill KA (2012) Markov processes follow from
594 the principle of maximum caliber. The Journal of Chemical Physics
595 136(6):064,108
- 596 Girardin V (2004) Entropy Maximization for Markov and Semi-Markov Pro-
597 cesses. Methodology and Computing in Applied Probability 6(1):109–127
- 598 Girardin V, Limnios N (2003) On the Entropy for Semi-Markov Processes.
599 Journal of Applied Probability 40(4):1060–1068
- 600 Haddad WM, Chellaboina V, Hui Q (2010) Nonnegative and Compartmental
601 Dynamical Systems. Princeton University Press
- 602 Harman CJ, Kim M (2014) An efficient tracer test for time-variable transit
603 time distributions in periodic hydrodynamic systems. Geophysical Research
604 Letters 41(5):1567–1575
- 605 Jacquez JA, Simon CP (1993) Qualitative theory of compartmental systems.
606 Siam Review 35(1):43–79
- 607 Jaynes ET (1957a) Information Theory and Statistical Mechanics. Physical
608 Review 106(4):620–630
- 609 Jaynes ET (1957b) Information Theory and Statistical Mechanics. ii. Physical
610 Review 108(2):171–190
- 611 Jaynes ET (1985) Macroscopic prediction. In: Complex Systems – Operational
612 Approaches in Neurobiology, Physics, and Computers, Springer, pp 254–269
- 613 Kloeden PE, Pötzsche C (2013) Nonautonomous Dynamical Systems in the
614 Life Sciences. Springer
- 615 Manzoni S, Porporato A (2009) Soil carbon and nitrogen mineralization: The-
616 ory and models across scales. Soil Biology and Biochemistry 41(7):1355–
617 1379, DOI 10.1016/j.soilbio.2009.02.031
- 618 Matis JH, Patten BC, White GC (1979) Compartmental Analysis of Ecosys-
619 tem Models, vol 10. International Co-operative Publishing House
- 620 Metzler H, Sierra CA (2018) Linear autonomous compartmental models as
621 continuous-time Markov chains: Transit-time and age distributions. Mathe-
622 matical Geosciences 50(1):1–34, DOI 10.1007/s11004-017-9690-1
- 623 Metzler H, Müller M, Sierra CA (2018) Transit-time and age distributions
624 for nonlinear time-dependent compartmental systems. Proceedings of the
625 National Academy of Sciences 115(6):1150–1155
- 626 Nash J (1957) The form of the instantaneous unit hydrograph. International
627 Association of Scientific Hydrology 3(45):114–121
- 628 Neuts MF (1981) Matrix-geometric solutions in stochastic models: An algo-
629 rithmic approach. The Johns Hopkins University Press
- 630 Norris JR (1997) Markov Chains. Cambridge University Press
- 631 Pesin YB (1977) Characteristic Lyapunov exponents and smooth ergodic the-
632 ory. Uspekhi Matematicheskikh Nauk 32(4):55–112
- 633 Pressé S, Ghosh K, Lee J, Dill KA (2013) Principles of maximum entropy
634 and maximum caliber in statistical physics. Reviews of Modern Physics
635 85(3):1115

- 636 Rasmussen M, Hastings A, Smith MJ, Agusto FB, Chen-Charpentier BM,
 637 Hoffman FM, Jiang J, Todd-Brown KEO, Wang Y, Wang YP, Luo Y (2016)
 638 Transit times and mean ages for nonautonomous and autonomous compartmental
 639 systems. *Journal of Mathematical Biology* 73(6-7):1379–1398
- 640 Rodhe H, Björkström A (1979) Some consequences of non-proportionality between
 641 fluxes and reservoir contents in natural systems. *Tellus* 31(3):269–278
- 642 Shannon CE, Weaver W (1949) *The Mathematical Theory of Communication*.
 643 The University of Illinois Press, Urbana
- 644 Sierra CA, Müller M (2015) A general mathematical framework for representing
 645 soil organic matter dynamics. *Ecological Monographs* 85(4):505–524,
 646 DOI 10.1890/15-0361.1
- 647 Sierra CA, Müller M, Metzler H, Manzoni S, Trumbore SE (2016) The muddle
 648 of ages, turnover, transit, and residence times in the carbon cycle. *Global
 649 Change Biology* in print, DOI 10.1111/gcb.13556
- 650 Thompson MV, Randerson JT (1999) Impulse response functions of terres-
 651 trial carbon cycle models: method and application. *Global Change Biology*
 652 5(4):371–394, DOI 10.1046/j.1365-2486.1999.00235.x
- 653 Trucco E (1956) A note on the information content of graphs. *Bulletin of
 654 Mathematical Biology* 18(2):129–135
- 655 Walter GG (1986) Size identifiability of compartmental models. *Mathematical
 656 Biosciences* 81(2):165–176
- 657 Walter GG, Contreras M (1999) *Compartmental Modeling with Networks*.
 658 Birkhäuser
- 659 Wang YP, Chen BC, Wieder WR, Leite M, Medlyn BE, Rasmussen M, Smith
 660 MJ, Agusto FB, Hoffman F, Luo YQ (2014) Oscillatory behavior of two
 661 nonlinear microbial models of soil carbon decomposition. *Biogeosciences*
 662 11(7):1817–1831, DOI 10.5194/bg-11-1817-2014

663 **A Basic ideas of Shannon information entropy**

664 We introduce basic concepts of information entropy along the lines of Cover and Thomas
 665 (2006). There are two concepts of entropy of a random variable, depending on whether the
 666 random variable has a discrete or a continuous distribution.

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

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

667 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

$$\mathbb{H}(Y_c) = - \int_{R_c} f(y) \log f(y) dy = -\mathbb{E} [\log f(Y_c)].$$

668 *Remark 1* Depending on the base of the logarithm, the unit of the entropy changes. For
 669 base 2, the unit is called bits and for the natural logarithm with base e , the unit is called
 670 nats. If not stated differently, we use the value e as logarithmic base, i.e., we use the natural
 671 logarithm.

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

The left panel of Figure 1 depicts the Shannon entropy of a Bernoulli random variable Y_d with $\mathbb{P}(Y_d = 1) = 1 - \mathbb{P}(Y_d = 0) = p$ with $p \in [0, 1]$. 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

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

684 The right panel of Figure 1 shows the differential entropy of an exponentially distributed
 685 random variable $Y_c \sim \text{Exp}(\lambda)$ with rate parameter $\lambda > 0$, probability density function
 686 $f(y) = \lambda e^{-\lambda y}$ 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 is, the likelier is an early exit of the particle, and the easier it is to predict the moment of exit. Hence, the differential entropy decreases with increasing λ . It is given by

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

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

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

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

688 **Definition 3** Let Y_1 and Y_2 be two discrete random variables with joint probability mass
 689 function p . Furthermore, let p_2 denote the probability mass function of Y_2 and denote by
 690 $p(y_1 | y_2)$ the conditional probability $\mathbb{P}(Y_1 = y_1 | Y_2 = y_2)$.

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

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

691 The joint entropy of two random variables is the entropy of one variable plus the con-
692 ditional entropy of the other. This is expressed in

$$\mathbb{H}(Y_1, Y_2) = \mathbb{H}(Y_2) + \mathbb{H}(Y_1 | Y_2). \quad (15)$$

693 Let Y_3 be a third discrete random variable. Then

$$\mathbb{H}(Y_1, Y_2 | Y_3) = \mathbb{H}(Y_1 | Y_3) + \mathbb{H}(Y_2 | Y_1, Y_3). \quad (16)$$

694 Let Y_1, Y_2, \dots, Y_n be discrete random variables. By repeated application of Eq. (15) and
695 Eq. (16), we obtain the *chain rule*

$$\mathbb{H}(Y_1, Y_2, \dots, Y_n) = \sum_{k=1}^n \mathbb{H}(Y_k | Y_{k-1}, \dots, Y_1). \quad (17)$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

708 if the limit exists.

709 **B Proves of the MaxEnt examples**

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

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

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

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

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

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

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

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

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

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

We rewrite the left hand side as

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

- 713 which leads to the constraints

$$C_{2,i} = \frac{1}{\|\mathbf{u}\|} \sum_{j=1, j \neq i}^d B_{ij} x_j^* - \frac{1}{\|\mathbf{u}\|} x_i^* \left(\sum_{k=1, k \neq i}^d B_{ki} + z_i \right) + \beta_i = 0, \quad i \in S. \quad (18)$$

- 714 The Lagrangian is now given by

$$L = \mathbb{H}(X) + \gamma_0 C_1 + \sum_{i=1}^d \gamma_i C_{2,i} \quad (19)$$

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

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

and

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

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

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

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

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

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

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

The case $\gamma_0 = 1$ has no solution \mathbf{v} since $e^{\gamma_i} > 0 > -1$. For $\gamma_0 \neq 1$ the matrix \mathbf{Y} has a nonzero determinant which makes the system uniquely solvable. For symmetry reasons, $\gamma_i = \gamma_j =: \gamma$ for all $i, j = 1, 2, \dots, d$. Consequently, for $i \neq j$, we get $B_{ij} = 1$, and by summing Eq. (18) over $i \in S$,

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

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

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

Uniqueness of this solution follows from its construction, we remain with showing maximality. To this end, we split the entropy into three parts, i.e., $\mathbb{H}(X) = H_1 + H_2 + H_3$ with

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

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, Section 5.3),

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

to rewrite the second term as

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

715 where T_E denotes the exponentially distributed sojourn time in E right before absorption.
 716 We see that H_2 becomes maximal if the knowledge of E contains no information about T_E .
 717 Hence, $z_j = z_i$ for $i, j \in S$. Since we need to ensure the systems' constraint on $\mathbb{E}[\mathcal{T}]$, we get
 718 $z_j = 1/\mathbb{E}[\mathcal{T}]$ for all $j \in S$.

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

$$H_3 = \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d 1 \cdot (1 - \log 1) = (d-1) \sum_{j=1}^d \mathbb{E}[O_j] = (d-1) \mathbb{E}[\mathcal{T}]$$

by Eq. (8). 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, i.e., $\mathbb{E}[\mathcal{T}(\varepsilon)] = \mathbb{E}[\mathcal{T}]$. The ε -disturbed H_3 is given by

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

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

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

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

Proposition B.2 Consider the set \mathcal{M}_2 of compartmental systems in equilibrium given by Eq. (1) with a predefined nonzero input vector \mathbf{u} and a predefined positive steady-state vector \mathbf{x}^* . The compartmental system $M_2^* = (\mathbf{u}, \mathbf{B}^*)$ with $\mathbf{B}^* = (B_{ij})_{i,j \in S}$ given by

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

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

Proof The mean transit time $\mathbb{E}[\mathcal{T}] = \|\mathbf{x}^*\|/\|\mathbf{u}\|$ of the system is fixed. Hence, the Lagrangian L is the same as in Eq. (19), and setting $\partial L/\partial B_{ij} = 0$ leads to

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

An interchange of the indices and summing the two equations gives

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

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

$$-\log z_j - \gamma_j = 0, \quad j \in S,$$

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

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

and

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