

Information Content and Maximum Entropy of Compartmental Systems in Equilibrium

Holger Metzler^{1,2,3,4,*}  and Carlos A. Sierra¹ ¹ Max Planck Institute for Biogeochemistry, Hans-Knöll-Str. 10, 07745 Jena, Germany; csierra@bgc-jena.mpg.de² Department of Geography, Ludwig-Maximilians-Universität Munich, Luisenstr. 37, 80333 Munich, Germany³ Department of Crop Production Ecology, Swedish University of Agricultural Sciences, Ulls väg 16, 756 51 Uppsala, Sweden⁴ Department of Forest Ecology and Management, Swedish University of Agricultural Sciences, Skogsmarksgränd 17, 901 83 Umeå, Sweden

* Correspondence: holger.metzler@lmu.de; Tel.: +49-89-218-06-548

Abstract

Mass-balanced compartmental systems defy classical deterministic entropy measures since both metric and topological entropy vanish in dissipative dynamics. By interpreting open compartmental systems as absorbing continuous-time Markov chains that describe the random journey of a single representative particle, we allow established information-theoretic principles to be applied to this particular type of deterministic dynamical system. In particular, path entropy quantifies the uncertainty of complete trajectories, while entropy rates measure the average uncertainty of instantaneous transitions. Using Shannon's information entropy, we derive closed-form expressions for these quantities in equilibrium and extend the maximum entropy principle (MaxEnt) to the problem of model selection in compartmental dynamics. This information-theoretic framework not only provides a systematic way to address equifinality but also reveals hidden structural properties of complex systems such as the global carbon cycle.

Keywords: information entropy; compartmental systems; equifinality; model identification; MaxEnt; reservoir models



Academic Editors: Jasper A. Vrugt, Anneli Guthke and Steven V. Weijss

Received: 31 August 2025

Revised: 10 October 2025

Accepted: 14 October 2025

Published: 21 October 2025

Citation: Metzler, H.; Sierra, C.A. Information Content and Maximum Entropy of Compartmental Systems in Equilibrium. *Entropy* **2025**, *27*, 1085. <https://doi.org/10.3390/e27101085>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

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 [1,2]. In principle, we are interested in selecting models that are parsimonious, i.e., models with the least degree of complexity for explaining certain patterns in nature [3]. 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 [4]. These metrics have been used extensively to characterize chaotic behavior in complex nonlinear systems [5], but they give trivial results for a large range of models used in the natural sciences.

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 [6–8].

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 that have been widely studied for systems in and out of equilibrium [9–14]. They help compare the behavior and quality of different models. Nevertheless, structurally very different models might show very similar ages and transit times and might represent a given measurement equally well. If we are in the position to select one of such models, which is the one to select? This equifinality problem can be resolved by the maximum entropy principle (MaxEnt) [15,16], a generic procedure to draw unbiased inferences from measurement or stochastic data [3,17].

In order to apply MaxEnt to compartmental systems, an 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 trajectories with slightly disturbed initial conditions do not diverge. Hence, by Pesin's theorem [18], 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 [19] provide a comprehensive overview of the history of graph entropy measures. Unfortunately, most such entropy measures are based on the number of vertices, vertex degree, edges, or degree sequence [20,21]. Thus, they concentrate 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 [22].

To bridge this gap, we interpret deterministic compartmental systems from a probabilistic viewpoint which allows us to apply the whole information theoretical toolbox to this important class of deterministic systems. As a first step in this direction, we compute the Shannon information entropy [23] of the continuous-time Markov chain that describes the random path of a single particle through the compartmental system [13] and introduce three non-vanishing 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.

The focus on a single particle gives our entropies microscopic system properties and consequently distinguishes our approach from the theory of maximum caliber (Max-Cal) [24,25], where path entropy is interpreted as a macroscopic system property of bulk material. Furthermore, our approach differs from the thermodynamic approach to entropy, which has been developed by other authors studying energy transfers and reversibility in thermodynamic systems [8,26–28]. While the probabilistic interpretation of thermodynamic entropy is related to the uncertainty of the location of a typical particle at a specific point in time, the newly introduced path entropy considers all locations of a typical particle at all times while it is part of the system. A first application of our information theoretical entropy concept to compartmental systems allows us to reveal hidden inherent properties of complex systems such as the carbon cycle, from the microbial to the global scale, e.g., allowing us to partly explain why there is a large diversity of soil carbon models, while there is more consensus on how to model carbon uptake by photosynthesis.

This article is organized as follows. First, we provide the fundamentals from information theory and dynamical systems theory that are necessary to introduce path entropy as the uncertainty of a single particle traveling through the system. Then, we mathematically derive the path entropy and introduce the entropy rates per unit time and per jump as uncertainty measures of the behavior of one typical particle. The focus on a single particle gives our entropies microscopic system properties and provides insights, where the macroscopic approach via topological and metric entropy fails. Then, we prove that the new entropy

rates of a finite particle path are indeed proper entropy rates of an associated stationary stochastic process, which guarantees that, in the process of model selection, there exists a unique optimal solution as long as the parameter set is convex. To assist in the interpretation of the newly introduced quantities from a system-wide point of view, we establish the link to the macroscopic system scale before we introduce the link between MaxEnt and structural model identification for compartmental systems. 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. Mathematical Background: Information Entropy 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.

2.1. Short Summary of Shannon Information Entropy

We introduce a few basic concepts of information entropy. Within the framework of this article, discrete entropies are usually associated with 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 path while it is in the system. See (Cover and Thomas [29] Sects. 2 and 8) for a more detailed introduction to Shannon's information entropy and differential entropy. Entropy rates for discrete- and continuous-time stochastic processes are introduced in (Cover and Thomas [29] Sect. 4) and Bad Dumitrescu [30].

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

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

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, i.e., 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 , i.e., 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, Shannon and differential entropy have one main difference. Shannon entropy is always nonnegative, whereas differential entropy can have negative values. While Shannon entropy is an absolute measure of information and makes sense in its own right, 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 Figure 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. 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 Figure 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, the more likely an early exit of the particle, 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 .

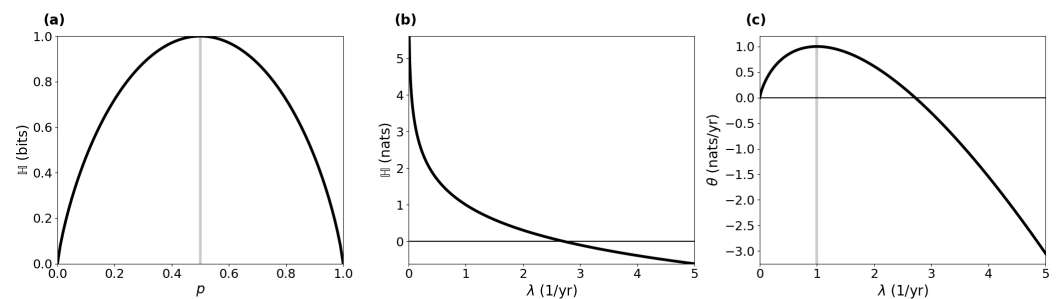


Figure 1. Shannon entropy of a Bernoulli distribution (a), differential entropy of an exponential distribution (b), and entropy rate of a Poisson process (c). Vertical gray lines indicate the parameter values leading to the highest entropy.

The *joint entropy* of two random variables Y_1 and Y_2 can be described as

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

where the *conditional entropy* $\mathbb{H}(Y_2 | Y_1)$ describes the uncertainty of Y_2 under the condition that Y_1 is known. The uncertainty of a stochastic process can be measured by its *entropy rate*, which describes the time density of the average information in the process. For a discrete-time stochastic process $Y = (Y_n)_{n \geq 1}$, it is defined as ([29] Sect. 4.2)

$$\theta(Y) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{H}(Y_1, Y_2, \dots, Y_n), \tag{5}$$

when the limit exists.

For instance, let $Z \sim \text{Poi}(\lambda)$ be a Poisson process with intensity rate $\lambda > 0$ describing the moments of occurrences of certain events. The interarrival times $Y = (Y_1, Y_2, \dots)$ of Z (the times between events) are $\text{Exp}(\lambda)$ -distributed and mutually independent. Hence, $\theta(Y) = 1 - \log \lambda$. If we rescale $\theta(Y)$ by the mean interarrival time $1/\lambda$, we obtain the entropy rate of Z being ([31] Sect. 3.3)

$$\theta(Z) = \theta(\text{Poi}(\lambda)) = \lambda (1 - \log \lambda). \tag{6}$$

This entropy rate increases with $\lambda \in [0, 1]$, reaches its maximum at 1, and then it decreases (Figure 1c). The maximum always occurs at $\lambda = 1$ independent of the unit of λ , because it

is based on the differential entropy of the exponential distribution, which is 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

The 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 [6,32]. Compartments are always well-mixed and usually also called *pools* or *boxes*. An autonomous compartmental system can be described by the d -dimensional linear ODE system

$$\frac{d}{dt} x(t) = Bx(t) + u, \quad t > 0, \tag{7}$$

with some nonnegative initial condition $x(0) = x^0 \in \mathbb{R}_+^d$. The nonnegative vector $x(t)$ describes the amount of material in the different compartments at time t , the nonnegative vector $u = (u_i)_{i=1,2,\dots,d} \in \mathbb{R}_+^d$ is the vector of external inputs to the compartments, and the compartmental matrix $B \in \mathbb{R}^{d \times d}$ describes the flux rates between the compartments and out of the system. The nonnegative off-diagonal value B_{ij} is the flux rate from compartment j to compartment i , the absolute value of the negative diagonal value B_{jj} is the total rate of fluxes out of compartment j , and the nonnegative column sum $z_j = -\sum_{i=1}^d B_{ij}$ 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*, i.e., all material that enters the system will eventually also leave it. Throughout this manuscript, we consider the open compartmental system (7) to have reached its unique steady-state or equilibrium compartment vector $x^* = -B^{-1}u$. This implies $\|r\| = \|u\|$, where $r = (r_i)_{i=1,2,\dots,d}$ 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 Equation (7) is fully characterized by u and B , and we denote it by $M := M(u, B)$.

2.3. The One-Particle Perspective

While Equation (7) 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 [13]. If $X_t \in S := \{1, 2, \dots, d\}$ 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 [33] on $\tilde{S} := S \cup \{d + 1\}$. Its initial distribution is given by $\tilde{\beta} = (\beta_1, \beta_2, \dots, \beta_d, 0)^T$, where $\beta := u/\|u\|$, 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 respective vector or matrix. The transition-rate matrix of X is given by

$$Q = \begin{pmatrix} B & \mathbf{0} \\ z^T & 0 \end{pmatrix}, \tag{8}$$

and thus,

$$\mathbb{P}(X_t = i) = (e^{tQ} \tilde{\beta})_i = \sum_{j=1}^d (e^{tQ})_{ij} \beta_j, \quad i \in \tilde{S} \tag{9}$$

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

$$\mathbb{P}(X_t = i | X_s = j) = (e^{(t-s)Q})_{ij}, \quad s \leq t, \quad i, j \in \tilde{S} \tag{10}$$

is the probability that X is in state i at time t , given it was in state j at a previous time s . Since the Markov chain X and the compartmental system in equilibrium given by Equation (7) are equivalent, we can write

$$M = M(\mathbf{u}, \mathbf{B}) = M(X). \tag{11}$$

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

$$\mathcal{P}(X) := ((Y_1 = X_0, T_1), (Y_2, T_2), \dots, (Y_{\mathcal{N}-1}, T_{\mathcal{N}-1}), Y_{\mathcal{N}} = d + 1), \tag{12}$$

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

$$\mathcal{N} := \inf \{n \in \mathbb{N} : Y_n = d + 1\} \tag{13}$$

denotes the first hitting time of the absorbing state $d + 1$ by the *embedded jump chain* $Y := (Y_n)_{n=1,2,\dots,\mathcal{N}}$ of X [33]. With $\lambda_j := -Q_{jj}$, the one-step transition probabilities of Y are given by, for $i, j \in \tilde{S}$,

$$P_{ij} := \mathbb{P}(Y_{n+1} = i | Y_n = j) = \begin{cases} 0, & i = j \text{ or } \lambda_j = 0, \\ Q_{ij} / \lambda_j, & \text{else.} \end{cases} \tag{14}$$

Let $P|_S = (P_{ij})_{i,j \in S}$ be the restriction of P to S . We can also write $P|_S = \mathbf{B}\mathbf{D}^{-1} + \mathbf{I}$, where $\mathbf{D} := \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d)$ is the diagonal matrix with the diagonal entries of \mathbf{B} , and \mathbf{I} denotes the identity matrix of appropriate dimension. Then, $\mathbf{M} := (\mathbf{I} - P|_S)^{-1}$ is the *fundamental matrix* of Y . The entry M_{ij} denotes the expected number of visits to compartment i given that the particle entered the system through compartment j . Consequently, the expected number of visits to compartment $i \in S$ is given by

$$\mathbb{E}[N_i] = \sum_{j=1}^d M_{ij} \beta_j = (\mathbf{M}\boldsymbol{\beta})_i = [(\mathbf{I} - P|_S)^{-1} \boldsymbol{\beta}]_i = (\mathbf{D}\mathbf{B}^{-1} \boldsymbol{\beta})_i = \frac{\lambda_i x_i^*}{\|\mathbf{u}\|}, \tag{15}$$

and the total expected number of jumps is given by

$$\mathbb{E}[\mathcal{N}] = \sum_{i=1}^d (\mathbf{M}\boldsymbol{\beta})_i + 1 = \sum_{i=1}^d \mathbb{E}[N_i] + 1, \tag{16}$$

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

$$\mathcal{T} := \inf \{t > 0 : X_t = d + 1\}. \tag{17}$$

The absorption time is phase-type distributed [34], $\mathcal{T} \sim \text{PH}(\boldsymbol{\beta}, \mathbf{B})$, with probability density function

$$f_{\mathcal{T}}(t) = \mathbf{z}^T e^{t\mathbf{B}} \boldsymbol{\beta}, \quad t \geq 0. \tag{18}$$

It can be shown ([13] Sect. 3.2) that the mean or expected value of \mathcal{T} equals the turnover time [12] of system (7) in equilibrium and is given by total stocks over total fluxes, i.e.,

$$\mathbb{E}[\mathcal{T}] = \frac{\|\mathbf{x}^*\|}{\|\mathbf{u}\|}. \tag{19}$$

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

$$\mathbb{1}_{\{A\}} = \begin{cases} 1, & A \text{ is true,} \\ 0, & \text{else,} \end{cases} \tag{20}$$

then $O_j := \sum_{k=1}^{\mathcal{N}-1} \mathbb{1}_{\{Y_k=j\}} T_k$ is the total time that the particle spends in compartment j . This time is called *occupation time* of j , and its mean is given by ([13] Sect. 3.3)

$$\mathbb{E}[O_j] = \frac{x_j^*}{\|\mathbf{u}\|}, \tag{21}$$

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

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 a summary of the desirable relations among them below:

- (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 particle’s 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 as to 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 $\mathcal{P} = \mathcal{P}(X)$ given by Equation (12) 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 \mathcal{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 \mathcal{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 $\mathcal{P}(X)$ of a single particle. Along the lines of Albert [35], 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 $\wp := \bigcup_{n=1}^\infty \wp_n$, we 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 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 $p \in \wp$ ([36] Chapter VI). Consequently, we can represent X by a finite-length path $\mathcal{P}(X) = ((Y_1, T_1), (Y_2, T_2), \dots, (Y_n, T_n), Y_{n+1})$ 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 $i \neq j$, denote by $N_{ij}(p)$ the total number of path's p 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*

$$f_{\mathcal{P}}(p) = \beta_{x_1} \left(\prod_{j=1}^d \prod_{i=1, i \neq j}^{d+1} (Q_{ij})^{N_{ij}(p)} \right) \prod_{j=1}^d e^{-\lambda_j R_j(p)}, \tag{22}$$

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

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

$$\begin{aligned} & \mathbb{P}((Y_1 = x_1, T_1 \leq t_1), \dots, (Y_n = x_n, T_n \leq t_n), Y_{n+1} = d + 1) \\ &= \mathbb{P}(Y_{n+1} = d + 1 \mid Y_n = x_n) \\ & \quad \cdot \prod_{k=2}^n \mathbb{P}(Y_k = x_k, T_k \leq t_k \mid Y_{k-1} = x_{k-1}) \mathbb{P}(Y_1 = x_1, T_1 \leq t_1) \\ &= P_{d+1, x_n} \left[\prod_{k=2}^n P_{x_k x_{k-1}} (1 - e^{-\lambda_{x_k} t_k}) \right] \beta_{x_1} (1 - e^{-\lambda_{x_1} t_1}) \\ &= \int_{\mathbb{T}_n} \beta_{x_1} \prod_{k=1}^n Q_{x_{k+1} x_k} e^{-\lambda_{x_k} \tau_k} d\tau_1 d\tau_2 \dots d\tau_n, \end{aligned} \tag{23}$$

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

$$f_{\mathcal{P}}(p) = \beta_{x_1} \prod_{k=1}^n Q_{x_{k+1} x_k} e^{-\lambda_{x_k} t_k}, \tag{24}$$

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

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

$$\begin{aligned} \prod_{k=1}^n e^{-\lambda_{x_k} t_k} &= \prod_{k=1}^n \prod_{j=1}^d \mathbb{1}_{\{x_k=j\}} e^{-\lambda_j t_k} = \prod_{j=1}^d e^{-\lambda_j \sum_{k=1}^n \mathbb{1}_{\{x_k=j\}} t_k} \\ &= \prod_{j=1}^d e^{-\lambda_j R_j(p)}. \end{aligned} \tag{25}$$

We make the according substitutions, and the proof is finished. \square

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

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

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\mathcal{P}) \\ &= -\sum_{i=1}^d \beta_i \log \beta_i \\ &\quad + \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j) \right]. \end{aligned} \tag{26}$$

Proof. Let X have the finite path representation

$$\mathcal{P} = \mathcal{P}(X) = ((Y_1, T_1), (Y_2, T_2), \dots, (Y_n, T_n), d + 1) \tag{27}$$

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

$$-\log f_{\mathcal{P}}(\mathcal{P}) = -\log \beta_{Y_1} - \sum_{j=1}^d \sum_{i=1, i \neq j}^{d+1} N_{ij}(\mathcal{P}) \log Q_{ij} + \sum_{j=1}^d \lambda_j R_j(\mathcal{P}). \tag{28}$$

We compute the expectation and get

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\mathcal{P}) = -\mathbb{E}[\log f_{\mathcal{P}}(\mathcal{P})] \\ &= -\mathbb{E}[\log \beta_{Y_1}] - \sum_{j=1}^d \sum_{i=1, i \neq j}^{d+1} \mathbb{E}[N_{ij}(\mathcal{P})] \log Q_{ij} + \sum_{j=1}^d \lambda_j \mathbb{E}[R_j(\mathcal{P})] \\ &= \mathbb{H}(Y_1) + \sum_{j=1}^d \lambda_j \mathbb{E}[R_j(\mathcal{P})] - \sum_{j=1}^d \sum_{i=1, i \neq j}^{d+1} \mathbb{E}[N_{ij}(\mathcal{P})] \log Q_{ij}. \end{aligned} \tag{29}$$

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 Equations (14) and (15),

$$\mathbb{E}[N_{ij}(\mathcal{P})] = \mathbb{E}[N_j(\mathcal{P})] P_{ij} = \begin{cases} \frac{x_j^*}{\|\mathbf{u}\|} B_{ij}, & i \leq d, \\ \frac{x_j^*}{\|\mathbf{u}\|} z_j, & i = d + 1. \end{cases} \tag{30}$$

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

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(Y_1) + \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \left[\left(\sum_{i=1, i \neq j}^d B_{ij} + z_j \right) \right. \\ &\quad \left. - \sum_{i=1, i \neq j}^d B_{ij} \log B_{ij} - z_j \log z_j \right] \\ &= - \sum_{i=1}^d \beta_i \log \beta_i + \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}) \right. \\ &\quad \left. + z_j (1 - \log z_j) \right]. \end{aligned} \tag{31}$$

□

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

$$\mathbb{H}(\boldsymbol{\beta}) := - \sum_{i=1}^d \beta_i \log \beta_i. \tag{32}$$

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

$$\mathbb{H}(X) = \mathbb{H}(\boldsymbol{\beta}) + \sum_{j=1}^d \mathbb{E}[O_j] \left(\sum_{i=1, i \neq j}^d \theta(\text{Poi}(B_{ij})) + \theta(\text{Poi}(z_j)) \right) \tag{33}$$

and

$$\begin{aligned} \mathbb{H}(X) &= \mathbb{H}(\boldsymbol{\beta}) \\ &\quad + \sum_{j=1}^d \mathbb{E}[N_j] \left(\mathbb{H}(\text{Exp}(\lambda_j)) + \mathbb{H}(P_{1,j}, P_{2,j}, \dots, P_{d,j}, P_{d+1,j}) \right), \end{aligned} \tag{34}$$

which can be rewritten as

$$\mathbb{H}(X) = \mathbb{H}(\boldsymbol{\beta}) + \sum_{j=1}^d \mathbb{E}[N_j] \mathbb{H}(P_{1,j}, P_{2,j}, \dots, P_{d,j}, P_{d+1,j}) \tag{35}$$

$$+ \sum_{j=1}^d \mathbb{E}[N_j] \mathbb{H}(\text{Exp}(\lambda_j)). \tag{36}$$

Proof. By virtue of Equation (33), we replace $x_j^*/\|\mathbf{u}\|$ by $\mathbb{E}[O_j]$ in Equation (26) and take into account that the entropy rate of a Poisson process with intensity rate λ equals $\lambda(1 - \log \lambda)$ to prove Equation (33). To prove Equation (34), we use Equation (15) to replace $x_j^*/\|\mathbf{u}\|$ in Equation (26) by $\mathbb{E}[N_j]/\lambda_j$ and obtain

$$\begin{aligned} \mathbb{H}(X) &= - \sum_{i=1}^d \beta_i \log \beta_i + \sum_{j=1}^d \mathbb{E}[N_j] (1 - \log \lambda_j) \\ &\quad + \sum_{j=1}^d \mathbb{E}[N_j] \left(- \sum_{i=1, i \neq j}^d \frac{B_{ij}}{\lambda_j} \log \frac{B_{ij}}{\lambda_j} - \frac{z_j}{\lambda_j} \log \frac{z_j}{\lambda_j} \right). \end{aligned} \tag{37}$$

Here, $(1 - \log \lambda_j)$ is the entropy of an exponential random variable with rate parameter λ_j . Using definition (14) of P_{ij} , we replace B_{ij}/λ_j by P_{ij} for $i \in S$ and z_j/λ_j by $P_{d+1,j}$ and finish the proof. □

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

Definition 1. *The path entropy of the compartmental system M in equilibrium given by Equation (7) with associated absorbing continuous-time Markov chain X and path $\mathcal{P} = \mathcal{P}(X)$ is defined by*

$$\mathbb{H}(\mathcal{P}) = \mathbb{H}(\mathcal{P}(X)) = \mathbb{H}(X). \tag{38}$$

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

$$\frac{d}{dt} x(t) = -\lambda x(t) + u, \quad t > 0 \tag{39}$$

and denote its associated path by \mathcal{P}_λ . The entropy of the initial distribution vanishes, and we obtain

$$\mathbb{H}(\mathcal{P}_\lambda) = \frac{x^*}{u} \lambda (1 - \log \lambda) = \frac{1}{\lambda} \lambda (1 - \log \lambda) = 1 - \log \lambda, \tag{40}$$

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

For a d -dimensional system, we can interpret $\mathbb{H}(\mathcal{P})$ as the entropy of a continuous-time process in the context of Equation (33) and as the entropy of a discrete-time process in the context of Equation (34). In both interpretations, the first term $\mathbb{H}(\beta) = \mathbb{H}(X_0) = \mathbb{H}(Y_1)$ represents the uncertainty of the first pool through which the particle enters the system. In the continuous-time interpretation, the uncertainty of the subsequent travel is the weighted average of the superposition of d Poisson processes describing the instantaneous uncertainty of possible jumps of the particle inside the system, $\theta(\text{Poi}(B_{ij}))$, and out of the system, $\theta(\text{Poi}(z_j))$, where the weights are the expected occupation times of the different compartments $j \in S$. 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 with each pool comprises the uncertainty of the length of the stay in the pool, $\mathbb{H}(\text{Exp}(\lambda_j))$, and the uncertainty of where to jump afterwards, $\mathbb{H}(\{P_{ij} : i \in \tilde{S}, j \in S, i \neq j\})$. Hence, in the context of Equation (34), we can separate the path entropy into a discrete part associated with the jump uncertainty given by Equation (35) and a continuous part associated with the sojourn time uncertainty given by Equation (36).

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

$$\theta(\mathcal{P}) = \frac{\mathbb{H}(\mathcal{P})}{\mathbb{E}[\mathcal{T}]} \tag{41}$$

and the *entropy rate per jump* by

$$\theta_J(\mathcal{P}) = \frac{\mathbb{H}(\mathcal{P})}{\mathbb{E}[\mathcal{N}]}. \tag{42}$$

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, it is the uncertainty entailed by the infinitesimal future, and for the entropy rate per jump, it is the uncertainty entailed by the next jump.

For these entropy rates to be useful in the process of model selection, it is important that they guarantee the existence of a unique maximum entropy model in case of a convex parameter space, which is not obvious from their definitions. The classical entropy rate of a stochastic process as defined in (Cover and Thomas [29] Sect. 4.1) has this property, and we prove in Appendix A that $\theta_J(\mathcal{P}) = \theta(Z)$, where the stationary process $Z = (Z_n)_{n \geq 1} = (\tilde{Y}_n, \tilde{T}_n)_{n \geq 1}$ on the space $(\tilde{S} \times \mathbb{R}_+)$ describes the infinite journey of a typical particle. It 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. By Equations (41) and (42), the average time between two jumps is $\mathbb{E}[\mathcal{T}] / \mathbb{E}[\mathcal{N}]$. If we divide the entropy rate per jump by it, we obtain the entropy rate per unit time. Hence,

$$\theta(\mathcal{P}) = \frac{\mathbb{E}[\mathcal{N}]}{\mathbb{E}[\mathcal{T}]} \theta(Z) \quad (43)$$

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

3.2. From Microscopic Particle Entropy to Macroscopic System Entropy

While the microscopic entropy measures provide direct insights into the uncertainties of the path of a single traveling particle, we can also scale them up to the macroscopic system scale. The combination of Equations (19), (41), and (43) inevitably leads to the following macroscopic definition.

Definition 2. *The system entropy of the compartmental system M in equilibrium given by Equation (7) with associated absorbing continuous-time Markov chain X and path $\mathcal{P} = \mathcal{P}(X)$ is defined by*

$$\mathbb{H}(M) = \|\mathbf{x}^*\| \theta(\mathcal{P}) = \|\mathbf{u}\| \mathbb{H}(\mathcal{P}) = \|\mathbf{u}\| \mathbb{E}[\mathcal{N}] \theta_J(\mathcal{P}). \quad (44)$$

Consequently, the system entropy can be interpreted in three ways: (1) as the cumulated mean instantaneous uncertainty of all particles currently in the system, (2) as the cumulated uncertainty of the entire future path of all particles currently entering the system, and (3) as the cumulated mean uncertainty of all future jumps of all particles currently entering the system.

3.3. 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 [17]. Its relationship to information theory and stochastics was established by Jaynes [15,16]. 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. For compartmental systems, data constraints could affect macroscopic quantities such as the stocks x^* , the input vector \mathbf{u} , the output rates z_j , or the mean transit time $\mathbb{E}[\mathcal{T}]$. The goal of this section is to transfer MaxEnt to compartmental systems in order to identify the compartmental system that best represents our state of knowledge in different situations and, at the same time, to get a better understanding of the previously 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 \mathcal{M}_1 of equilibrium compartmental systems (7) 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 $M_1^* \in \mathcal{M}_1$ with the path $\mathcal{P}_1^* := \mathcal{P}(M_1^*)$ such that the path entropy $\mathbb{H}(\mathcal{P}_1^*)$ or, equivalently, the entropy rate per unit time $\theta(\mathcal{P}_1^*)$ is maximized. We can show (see Proposition A2) that the compartmental system $M_1^* = M(\mathbf{u}, \mathbf{B})$ with

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

where $\lambda = d - 1 + 1/\mathbb{E}[\mathcal{T}]$, is the maximum entropy model in \mathcal{M}_1 . In the special case $d = 1$ for a one-dimensional compartmental system, we obtain $\mathbf{B} = -1/\mathbb{E}[\mathcal{T}]$. Since, in this case, $\mathcal{T} \sim \text{Exp}(-\mathbf{B})$, 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 ([29] Example 12.2.5).

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

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

is the maximum entropy model in \mathcal{M}_2 (see Proposition A3).

3.4. Structural Model Identification Assisted by MaxEnt

Suppose that 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 about the number of compartments that 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 (7) 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, i.e., the transfer function of the system, and identify the most dominant frequencies. The impulse response or the transfer function might be possible to obtain by tracer experiments [6,37].

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 [6] Chapter 16), the *structural identification problem* of linear autonomous systems is described as follows. Suppose that we are interested in determining a d -dimensional system of form (7).

We are interested in sending an impulse to the system at time $t = 0$ and analyzing its further behavior. To that end, we rewrite the system as

$$\begin{aligned} \frac{d}{dt} \mathbf{x}(t) &= \mathbf{B} \mathbf{x}(t) + \mathbf{A} \mathbf{u}, \quad t \geq 0, \\ \mathbf{x}(0) &= \mathbf{0} \in \mathbb{R}^d, \\ \mathbf{y}(t) &= \mathbf{C} \mathbf{x}(t), \quad t \geq 0. \end{aligned} \tag{47}$$

Note that the roles of \mathbf{A} and \mathbf{B} are interchanged here with respect to Anderson [6]. 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 [38] 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*

$$\Psi(t) := \mathbf{C} e^{t\mathbf{B}} \mathbf{A}, \quad t \geq 0, \tag{48}$$

or in the *transfer function*

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

which is the Laplace transform matrix of Ψ . Consequently, all identifiable parameters of \mathbf{A} , \mathbf{B} , and \mathbf{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 \mathbf{A} , \mathbf{B} , and \mathbf{C} . We call system (47) *identifiable* if this nonlinear system of equations has a unique solution $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ for given Ψ or $\widehat{\Psi}$. Otherwise, the system is called *non-identifiable*. Usually, the matrices \mathbf{A} and \mathbf{C} are already known from the experiment setup. What remains is to identify the compartmental matrix \mathbf{B} , and this can be achieved by MaxEnt.

4. Application to Particular Systems


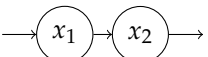
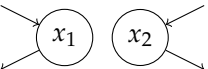
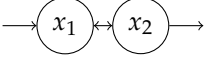
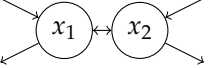

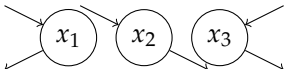
First, we apply the presented theory to some equilibrium compartmental models with very simple structure in order to grasp the new entropy concepts. Then, we compute entropy quantities for two carbon-cycle models in dependence on environmental and biochemical parameters. Finally, we apply MaxEnt to solve an equifinality problem in model selection as an example of 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 larger 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 not have the highest entropy rate per unit time or 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.

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 $\mathbb{E}[\mathcal{N}]$, entropy rate per unit time θ , mean transit time $\mathbb{E}[\mathcal{T}]$, and path entropy $\mathbb{H}(\mathcal{P})$. Underlined numbers are the highest values per column.

Structure	$\frac{d}{dt}x(t)$	θ_J	$\mathbb{E}[\mathcal{N}]$	θ	$\mathbb{E}[\mathcal{T}]$	$\mathbb{H}(\mathcal{P})$
	$-\lambda x + 1$	$0.5 (1 - \log \lambda)$	2.00	$\lambda(1 - \log \lambda)$	$1/\lambda$	$1 - \log \lambda$
	$\begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$	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	<u>4.00</u>	<u>5.39</u>
	$\begin{pmatrix} -1 & 1/2 \\ 1/2 & -1 \end{pmatrix} x + \begin{pmatrix} 1 \\ 1 \end{pmatrix}$	<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

4.2. A Linear Autonomous Global Carbon-Cycle Model

We consider the global carbon-cycle model introduced by Emanuel et al. [39] (Figure 2).

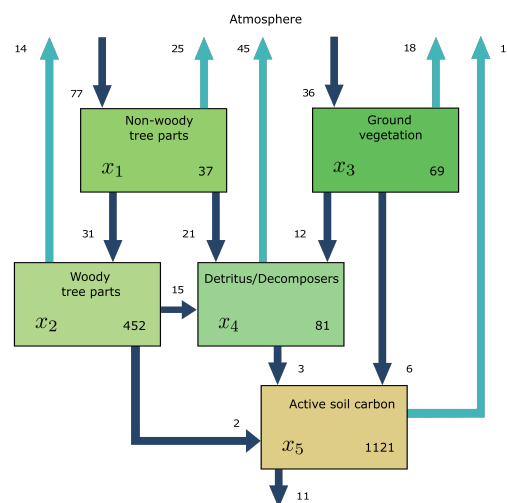


Figure 2. Schematic of the linear autonomous global carbon-cycle model in steady state, introduced by Emanuel et al. [39].

The model comprises five compartments: non-woody tree parts $x_1 = 37$ PgC, woody tree parts $x_2 = 452$ PgC, ground vegetation $x_3 = 69$ PgC, detritus/decomposers $x_4 = 81$ PgC, and active soil carbon $x_5 = 1121$ PgC. We introduce an environmental rate modifier ζ , which controls the speed at which carbon is cycled in all compartments. If

$\zeta > 1$, carbon is cycled faster in all compartments, simulating the effect of the global surface temperature increase [40]. For a given ζ , the equilibrium model $M_\zeta = M(\mathbf{u}, B_\zeta)$ is given by

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

and

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

where the numbers are chosen as in Thompson and Randerson [41]. The input vector is expressed in units of petagrams of carbon per year (PgC yr^{-1}) and the fractional transfer coefficients in units of per year (yr^{-1}). Because B_ζ is a lower triangular matrix, the model contains no feedbacks. For every value of ζ , the system has a different steady state (Figure 3a). The higher the value of ζ , the faster the system, which makes the mean transit time (Figure 3b) decrease, and because of shorter paths, the path entropy (Figure 3d) also decreases. Since ζ has no impact on the structure of the model, the mean number of jumps (Figure 3c) remains unaffected. This can also be seen from the solid line marked by squares in (Figure 3d). It represents the part of the path entropy related to jump-associated uncertainties (Equation (35)). The solid line marked by circles represents the part of the path entropy related to sojourn-associated uncertainties (Equation (36)), which as a weighted average of one-pool entropies decreases similarly to the entropy of an exponential distribution with an increasing rate parameter λ (Figure 1b). The two parts together constitute the path entropy as represented by the unmarked solid line.

The entropy rate per unit time (Figure 3e) increases until $\zeta \approx 6$ and decreases afterwards, because with increasing system speed the decreasing uncertainty associated with sojourn times increasingly dominates the uncertainty associated with jumps. While the uncertainty associated with jumps averaged over the path length increases, because the total jump uncertainty is constant (see solid line marked with squares in Figure 3d), and the mean path length decreases (Figure 3b), the sojourn-associated uncertainty decreases with the increasing system speed for $\zeta > 6$, similar to the entropy rate of a Poisson process with intensity rate $\lambda > 1$ (see Figure 1c). The entropy rate per jump (Figure 3f) decreases with increasing ζ , because the path entropy of the system decreases.

Dashed lines in Figure 3d–f show the respective entropy values for a one-pool system $M_\lambda = M((77 + 36) \text{ PgC yr}^{-1}, -\lambda)$ with the same mean transit time, i.e., $\lambda^{-1} = \mathbb{E}[\mathcal{T}_\zeta]$. The solid and dashed lines intersect at $\zeta \approx 4.31$ in Figure 3d,e. Before this break-even point, the path of this multiple-pool model is harder to predict than the path (i.e., the exit time of the particle) of a one-pool model with the same mean transit time. After this point of breaking even, the path of the 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.

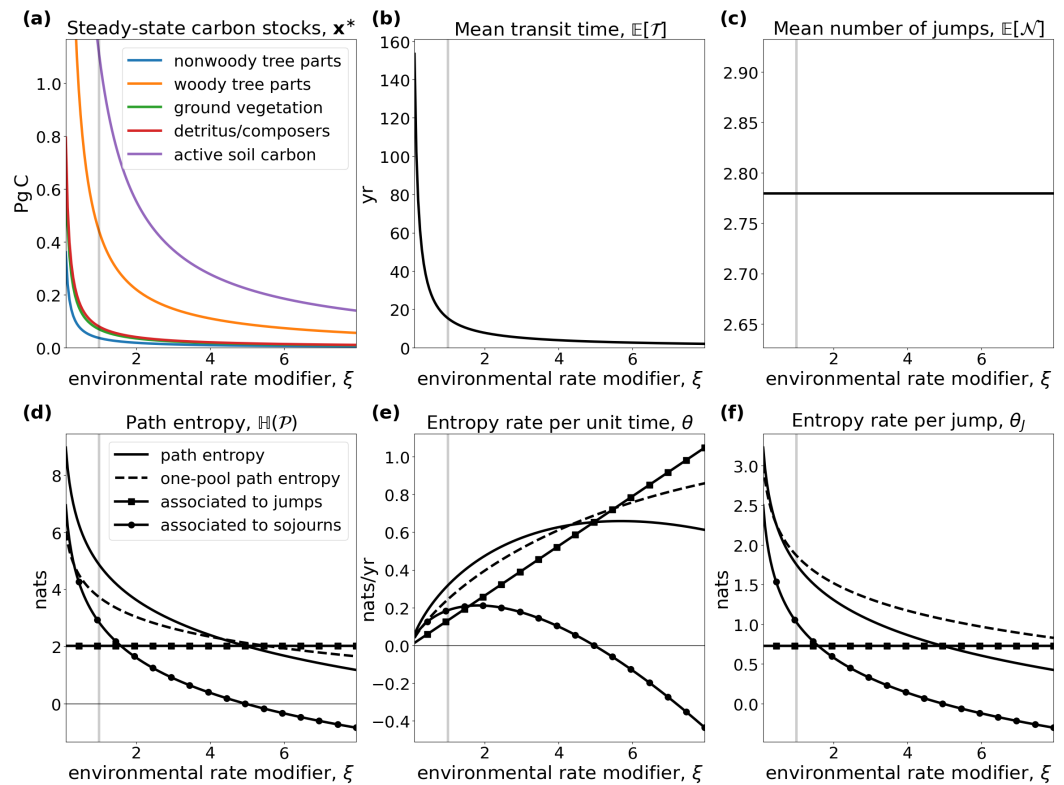


Figure 3. (a) Equilibrium carbon stocks and (b–f) entropy related quantities of the global carbon-cycle model, introduced by Emanuel et al. [39], in dependence on the environmental rate coefficient ξ . Vertical gray lines show $\xi = 1$, the original speed of the model.

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 only becomes 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 $M_\varepsilon = M(\mathbf{u}, B_\varepsilon)$, described by Wang et al. [42], which is used to represent the dynamics of microbes and carbon substrates in soils (Figure 4). Its ODE system is given by

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

where $\mathbf{x}(t) = (C_s, C_b)^T(t)$. We denote by C_s and C_b substrate organic carbon and soil microbial biomass carbon (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 ($\text{gC m}^{-2} \text{yr}^{-1}$), 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.

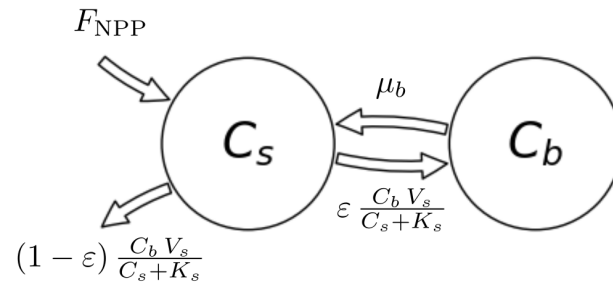


Figure 4. Scheme of the nonlinear autonomous carbon-cycle model, introduced by Wang et al. [42], with two compartments: substrate organic carbon (C_s) and microbial biomass (C_b).

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

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

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

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

From Wang et al. [42], we take the parameter values $\mu_b = 4.38 \text{ yr}^{-1}$, $F_{\text{NPP}} = 345.00 \text{ gC m}^{-2} \text{ yr}^{-1}$, and $K_s = 53,954.83 \text{ gC m}^{-2}$. Since the description of V_s is missing in the original publication, we let it be equal to 59.13 yr^{-1} to approximately meet the given steady-state contents $C_s^* = 12,650.00 \text{ gC m}^{-2}$ and $C_b^* = 50.36 \text{ gC m}^{-2}$ 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 \mathbf{B} is not triangular. For every value of ε , the system has a different steady state (Figure 5a). 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 Equation (54), the system speed increases, and the mean transit time goes down (Figure 5b) 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 (Figure 5c), making the entropy rate per jump decrease (Figure 5f). 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 Figure 5f). 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 (Figure 1c), which corresponds to $\varepsilon \approx 0.926$. This is also reflected in the entropy rate per unit time (Figure 5e). The maximum does not exactly occur at $\varepsilon = 0.926$, because the time 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 in the rate $-B_{11}$ of soil organic carbon turnover.

Considering a one-pool system $M_\lambda = M(345.00 \text{ gC m}^{-2} \text{ yr}^{-1}, -1/\mathbb{E}[\mathcal{T}_\varepsilon])$ with the same mean transit time, we 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 Figure 5d–f).

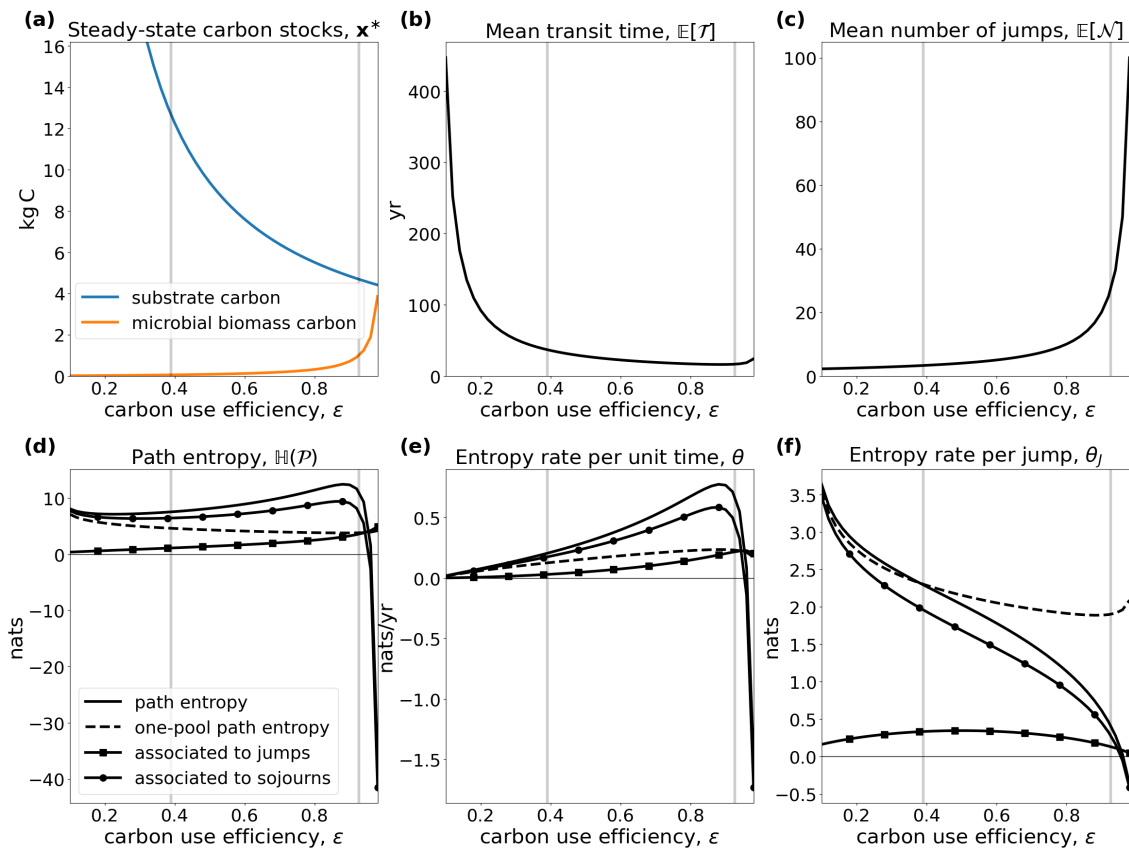


Figure 5. (a) Equilibrium carbon stocks and (b–f) entropy related quantities of the global carbon-cycle model, introduced by Wang et al. [42], in dependence on the microbial carbon use efficiency ϵ . The left vertical gray lines show $\epsilon = 0.39$, the original carbon use efficiency of the model, the right ones at $\epsilon = 0.926$ show the carbon use efficiency value with the maximum entropy rate of the Poisson process associated with C_5 .

4.4. Model Identification via Maxent

The following example is inspired by (Anderson [6] Example 16 C). It shows how MaxEnt can help make a decision about which model to use if not all parameters can be uniquely determined from the transfer function $\hat{\Psi}$. We are interested in determining the entries of the compartmental matrix B belonging to the two-dimensional equilibrium compartmental system

$$\frac{d}{dt} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (t) = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} (t) + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{gC yr}^{-1}, \quad t > 0. \tag{55}$$

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

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

where

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

We assume that $\hat{\Psi}$ is known from measurements, i.e., 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 non-identifiable and there remains a class \mathcal{M} of models which all satisfy Equation (57). Which model out of \mathcal{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 $M^* = M(\mathbf{u}, \mathbf{B}^*)$ such that

$$M^* = \arg \max_{M \in \mathcal{M}} \theta(\mathcal{P}(M)). \quad (58)$$

Maximizing the entropy rate per unit time here leads to a feasible optimization problem, whereas maximization of the path entropy by slowing down the model and indefinitely increasing its mean transit time and its path entropy would lead to an unbounded optimization problem. The parameter space associated with \mathcal{M} is given by

$$\{\mathbf{p} = (B_{12}, B_{21}, z_1, z_2) \in \mathbb{R}_+^4 : \mathbf{p} \text{ satisfies Equation (57)}\}, \quad (59)$$

which is not guaranteed to be convex in general. Consequently, by fundamental principles from mathematical optimization theory, the 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 $\gamma_1 = 3 \text{ yr}^{-1}$, $\gamma_2 = 5 \text{ yr}^{-1}$, and $\gamma_3 = 4 \text{ yr}^{-1}$. Since convexity of the parameter space is not guaranteed, local optimality does not guarantee global optimality. Hence, we run local optimizations 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. Even though we cannot rigorously prove that our global maximum candidate $M_{\max} = M(\mathbf{u}, \mathbf{B}_{\max})$, as represented by the red dot in Figure 6 with

$$\mathbf{B}_{\max} \approx \begin{pmatrix} -2.723 & 1.821 \\ 1.098 & -2.277 \end{pmatrix} \text{ yr}^{-1} \quad (60)$$

and $\theta_{\max} \approx 1.916$, is a global maximum, we can clearly see that it is a good candidate. Increasing the distance of the local maximum parameters (Figure 6a) and mean transit time (Figure 6b) from the global maximum candidate leads to a decrease in the entropy rate per unit time. 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 (Figure 6c). Finally, the global maximum candidate for the entropy rate per unit time does not maximize the path entropy (Figure 6d).

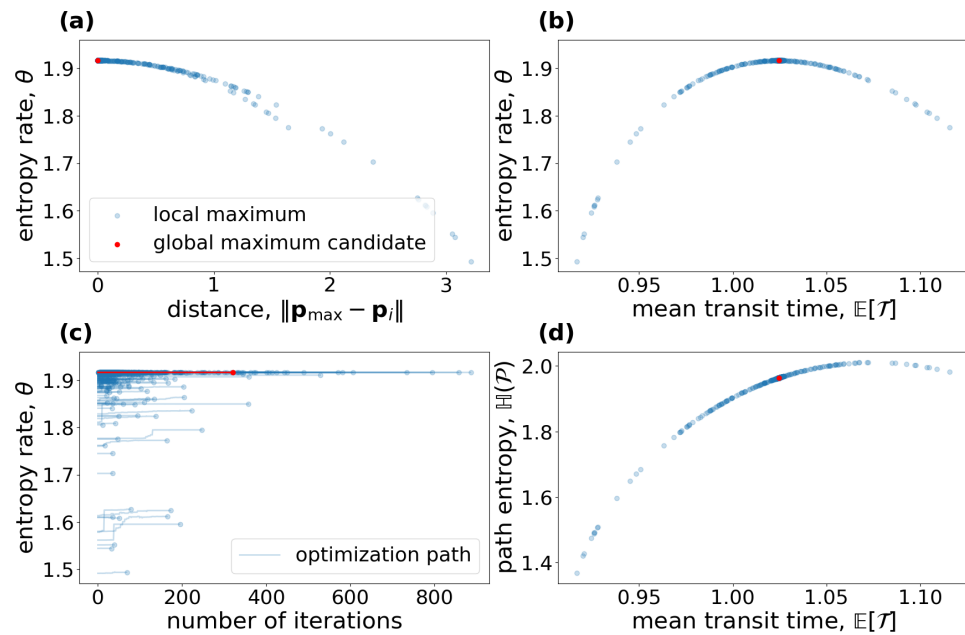


Figure 6. Local maximizations of θ over a grid on a subspace of the parameter space. For better visibility, we randomly chose 1000 grid points for the plot. Blue dots show local maxima found during the global maximization procedure starting on the grid. The red dot is associated with the global maximum candidate M_{\max} . (a) Entropy rate per unit time versus l_1 -distance of 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 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 $\mathbb{H}(\mathcal{P})$ the 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) [24,25]. We maximize here the Shannon information 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. [17], 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. We derive macroscopic system entropy by multiplying microscopic entropy quantities (e.g., path entropy, entropy rate per unit time) with the associated macroscopic system quantities (e.g., total system content, total input amount).

In the field of soil carbon cycle modeling, Ågren [43] 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 Section 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 instead of the path entropy, because, by slowing the system down more and more, the path entropy can potentially be increased indefinitely, and a maximum path entropy model does not exist. The decision to maximize a rate can potentially also be justified by a given macroscopic restriction on the stock sizes, which would increase indefinitely with indefinitely increasing path entropy by slowing down the system.

By virtue of its very mathematical definition (Equation (1)), entropy is maximized when the system's symmetry is maximized. This is indicated by the Bernoulli entropy (Figure 1a) 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. 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 $z_j = 1/\mathbb{E}[\mathcal{T}]$ of leaving the system are chosen such that the mean transit time constraint is fulfilled. In Example 2, the symmetry is broken by the additional restriction of a given steady-state vector. Consequently, $\mathbb{H}(\mathcal{P}_2^*) \leq \mathbb{H}(\mathcal{P}_1^*)$.

When we compute entropy values for actual carbon-cycle models (Sections 4.2 and 4.3), we note that environmental or eco-physiological factors might impact model entropies. For example, higher global surface temperatures are likely to induce a higher global carbon-cycle system speed ($1 \leq \xi \ll 6$). 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, i.e., 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 number 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 (Sections 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, and heterogeneity increases the uncertainty of a particle's sequence of visited compartments.

These simple insights allow us to understand modeling issues in a broader sense. For instance, path entropies support the understanding of differences in the diversity of modeling approaches and predictions for carbon uptake and transfers to soils in terrestrial ecosystems. Both photosynthesis [44] and soil carbon turnover [45] are modeled by many different approaches. However, in ecosystem models, photosynthesis is almost exclusively represented [46] by the Farquhar model [47], while soil carbon dynamics are represented by a great variety of models with very different structures [48]. The latter leads to large variations in the prediction of future land carbon uptake [48,49]. A comparison of carbon simulations from eleven modeling centers showed that across models, global soil carbon varied more than twice as much as global net primary productivity [50]. Carbon dynamics

in leaves are relatively fast, and the fate of this carbon is highly predictable; it is either used to fuel the metabolic activity of leaf cells or allocated to storage reserves or woody tissue [51]. In contrast, the fate of carbon entering soils is much less predictable, with a large range of potential metabolic pathways through microbial food webs or potential physico-chemical interactions with the mineral surfaces in the soil matrix that occur at longer timescales [52]. 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 possesses high inherent uncertainty, which hints at a theoretical limit that cannot be overcome by any model.

The example of model identification by MaxEnt in Section 4.4 shows a major difference from the more artificial previous maximum entropy examples. The given constraints do not tell us enough about the structure of the model class \mathcal{M} to ensure that an identified local maximum is also a global maximum. Owing to the nonlinear restrictions on the parameters in Equation (57), the parameter space is probably not convex. Hence, local maxima are not guaranteed to be also globally optimal. The small 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 optimization methods suitable for the particular problem at hand should then be applied. However, since the newly introduced entropy measures are proper entropies, and in the case of a convex parameter space, the existence of a unique global optimum is guaranteed.

6. Conclusions

A probabilistic approach to mass-balanced deterministic systems allows basic information theoretical principles to compute the uncertainty of a wide range of models representing complex processes in nature—a task at which classical deterministic theories fail. The information content of autonomous compartmental systems in equilibrium can be assessed by the entropy of the path of particles traveling through 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 occupied compartments, (3) as a discrete sequence of pairs consisting of visited compartments and associated sojourn times. Based on these three 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 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 can be achieved 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, which was so far applied only to the inhomogeneous embedded jump chain, as by Ge et al. [53], to incorporate sojourn times in different compartments as well.

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 based on Shannon information theory in both models and observations, we can potentially advance toward better methods for model selection applying the maximum entropy principle.

Author Contributions: H.M.: Writing—review and editing, Writing—original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. C.A.S.: Writing—review and editing, Validation, Funding acquisition, Supervision, Project administration, Conceptualization. All authors have read and agreed to the published version of the manuscript.

Funding: Open Access funding provided by the Max Planck Society. This research was funded by 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.

Data Availability Statement: The Python (version 3.12.2) code to reproduce the figures used in the manuscript is provided in the static repository <https://doi.org/10.5281/zenodo.17396760>.

Acknowledgments: H.M. acknowledges the support of Giulia Vico at the Swedish University of Agricultural Sciences in Uppsala, Oskar Franklin at the International Institute for Applied Systems Analysis, and Julia Pongratz at the Ludwig-Maximilians-Universität Munich for their support during the long-standing processes of transforming pure thought material into a publishable manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. The Stationary Process Z

We prove that the entropy rate per jump of a single traveling particle is a proper entropy rate in the sense of the definition as given in Equation (5). Let $Z = (Z_n)_{n \geq 1} = (\tilde{Y}_n, \tilde{T}_n)_{n \geq 1}$ on the space $(\tilde{S} \times \mathbb{R}_+)$ describe the infinite journey of a typical particle. It 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, defined by the transition probabilities $\tilde{P}_{ij}(t) = \mathbb{P}(\tilde{Y}_{n+1} = i, \tilde{T}_{n+1} \leq t \mid \tilde{Y}_n = j)$ given by

$$\tilde{P}_{ij}(t) = \begin{cases} 0, & i = j, \\ B_{ij} \lambda_j^{-1} (1 - e^{-\lambda_i t}), & i, j \leq d, i \neq j, \\ z_j \lambda_j^{-1}, & i = d + 1, j \leq d, \\ \beta_i (1 - e^{-\lambda_i t}), & i \leq d, j = d + 1, \end{cases} \tag{A1}$$

and initial (stationary) distribution

$$\pi_j(t) = \frac{1}{\mathbb{E}[\mathcal{N}]} \cdot \begin{cases} \mathbb{E}[N_j] (1 - e^{-\lambda_j t}), & j \leq d, \\ 1, & j = d + 1. \end{cases} \tag{A2}$$

Proposition A1. *The entropy rate per jump, $\theta_J(\mathcal{P})$, equals the entropy rate of the stationary process Z.*

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

$$\begin{aligned}
 \mathbb{P}(\tilde{Y}_2 = i, \tilde{T}_2 \leq t) &= \sum_{j=1}^d \mathbb{P}(\tilde{Y}_2 = i, \tilde{T}_2 \leq t \mid \tilde{Y}_1 = j) \mathbb{P}(\tilde{Y}_1 = j) \\
 &= \sum_{j=1}^d \tilde{P}_{d+1,j}(t) \pi_j \\
 &= \sum_{j=1}^d \frac{z_j}{\lambda_j} \frac{\mathbb{E}[N_j]}{\mathbb{E}[\mathcal{N}]}.
 \end{aligned}
 \tag{A3}$$

By Equation (15), $r_j = z_j x_j^*$, and $\|r\| = \|u\|$, we get

$$\mathbb{P}(\tilde{Y}_2 = i, \tilde{T}_2 \leq t) = \frac{1}{\mathbb{E}[\mathcal{N}]} \sum_{j=1}^d \frac{z_j}{\lambda_j} \frac{\lambda_j x_j^*}{\|u\|} = \frac{1}{\mathbb{E}[\mathcal{N}]} \frac{z^T x^*}{\|u\|} = \pi_{d+1}(t).
 \tag{A4}$$

Now let $i \leq d$. Then,

$$\begin{aligned}
 \mathbb{P}(\tilde{Y}_2 = i, \tilde{T}_2 \leq t) &= \sum_{j=1, j \neq i}^d \frac{B_{ij}}{\lambda_j} (1 - e^{-\lambda_i t}) \frac{\mathbb{E}[N_j]}{\mathbb{E}[\mathcal{N}]} + \beta_i (1 - e^{-\lambda_i t}) \frac{1}{\mathbb{E}[\mathcal{N}]} \\
 &= \frac{1}{\mathbb{E}[\mathcal{N}]} \left[\sum_{i=1, i \neq j}^d \frac{B_{ij} x_j^*}{\|u\|} + \beta_i \right] (1 - e^{-\lambda_i t}) \\
 &= \frac{1}{\mathbb{E}[\mathcal{N}]} \left[\frac{(\mathbf{B} x^*)_i}{\|u\|} - \frac{B_{ii} x_i^*}{\|u\|} + \beta_i \right] (1 - e^{-\lambda_i t}) \\
 &= \frac{1}{\mathbb{E}[\mathcal{N}]} \left[-\frac{u_i}{\|u\|} + \frac{\lambda_i x_i^*}{\|u\|} + \beta_i \right] (1 - e^{-\lambda_i t}) \\
 &= \frac{1}{\mathbb{E}[\mathcal{N}]} \mathbb{E}[N_i] (1 - e^{-\lambda_i t}) \\
 &= \pi_i(t).
 \end{aligned}
 \tag{A5}$$

Step 2. Since Z is stationary, by (Cover and Thomas [29] Theorem 4.2.1), its entropy rate given by

$$\theta(Z) = \lim_{n \rightarrow \infty} \mathbb{H}(Z_{n+1} \mid Z_n) = \mathbb{H}(Z_2 \mid Z_1),
 \tag{A6}$$

which computes to

$$\begin{aligned}
 \theta(Z) &= \mathbb{H}((\tilde{Y}_2, \tilde{T}_2) \mid (\tilde{Y}_1, \tilde{T}_1)) = \mathbb{H}((\tilde{Y}_2, \tilde{T}_2) \mid \tilde{Y}_1) = \mathbb{H}(\tilde{T}_2 \mid \tilde{Y}_2, \tilde{Y}_1) + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1) \\
 &= \mathbb{H}(\tilde{T}_2 \mid \tilde{Y}_2) + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1).
 \end{aligned}
 \tag{A7}$$

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

$$\begin{aligned}
 \theta(Z) &= \mathbb{H}(\tilde{T}_1 \mid \tilde{Y}_1) + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1) \\
 &= \sum_{j=1}^{d+1} \pi_j \left[\mathbb{H}(\tilde{T}_1 \mid \tilde{Y}_1 = j) + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1 = j) \right], \\
 &= \frac{1}{\mathbb{E}[\mathcal{N}]} \left(\sum_{j=1}^d \mathbb{E}[N_j] \left[\mathbb{H}(\tilde{T}_1 \mid \tilde{Y}_1 = j) + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1 = j) \right] + \mathbb{H}(\tilde{Y}_2 \mid \tilde{Y}_1 = d+1) \right),
 \end{aligned}
 \tag{A8}$$

which together with Equation (34) finishes the proof. \square

Appendix B. Proofs of the MaxEnt Examples

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

$$\begin{aligned} \mathbb{H}(\mathcal{P}(M)) &= \mathbb{H}(X) \\ &= - \sum_{i=1}^d \beta_i \log \beta_i + \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \left[\sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j) \right]. \end{aligned} \tag{A9}$$

In order to obtain maximum entropy models under simple constraints, we now adapt ideas of [54].

Proposition A2. Consider the set \mathcal{M}_1 of compartmental systems in equilibrium given by Equation (7) 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 $M_1^* = M(\mathbf{u}, \mathbf{B}^*)$ with

$$\mathbf{B}^* = \begin{pmatrix} -\lambda & 1 & \dots & 1 \\ 1 & -\lambda & 1 \dots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \dots & 1 & -\lambda \end{pmatrix}, \tag{A10}$$

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

$$C_1 = \frac{1}{\|\mathbf{u}\|} \sum_{j=1}^d x_j^* - \mathbb{E}[\mathcal{T}] = 0. \tag{A11}$$

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

$$\frac{1}{\|\mathbf{u}\|} (\mathbf{B} \mathbf{x}^*)_i = -\beta_i, \quad i = 1, 2, \dots, d. \tag{A12}$$

We rewrite the left hand side as

$$\begin{aligned} \frac{1}{\|\mathbf{u}\|} (\mathbf{B} \mathbf{x}^*)_i &= \frac{1}{\|\mathbf{u}\|} \sum_{j=1}^d B_{ij} x_j^* = \frac{1}{\|\mathbf{u}\|} \left(\sum_{j=1, j \neq i}^d B_{ij} x_j^* + B_{ii} x_i^* \right) \\ &= \frac{1}{\|\mathbf{u}\|} \sum_{j=1, j \neq i}^d B_{ij} x_j^* - \frac{1}{\|\mathbf{u}\|} x_i^* \left(\sum_{k=1, k \neq i}^d B_{ki} + z_i \right), \end{aligned} \tag{A13}$$

which leads to the constraints

$$C_{2,i} = \frac{1}{\|\mathbf{u}\|} \sum_{j=1, j \neq i}^d B_{ij} x_j^* - \frac{1}{\|\mathbf{u}\|} x_i^* \left(\sum_{k=1, k \neq i}^d B_{ki} + z_i \right) + \beta_i = 0, \quad i \in S. \tag{A14}$$

The Lagrangian is now given by

$$L = \mathbb{H}(X) + \gamma_0 C_1 + \sum_{i=1}^d \gamma_i C_{2,i} \tag{A15}$$

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

$$\begin{aligned} \|\mathbf{u}\| \frac{\partial}{\partial B_{ij}} L &= -x_j^* \log B_{ij} + \gamma_i x_j^* - \gamma_j x_j^*, \\ \|\mathbf{u}\| \frac{\partial}{\partial z_j} L &= -x_j^* \log z_j - \gamma_j x_j^*, \end{aligned} \tag{A16}$$

and

$$\begin{aligned} \|\mathbf{u}\| \frac{\partial}{\partial x_j^*} L &= \sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}) + z_j (1 - \log z_j) \\ &+ \gamma_0 + \sum_{i=1, i \neq j}^d \gamma_i B_{ij} - \gamma_j \left(\sum_{k=1, k \neq j}^d B_{kj} + z_j \right), \end{aligned} \tag{A17}$$

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

$$\begin{aligned} 0 &= \sum_{i=1, i \neq j}^d e^{\gamma_i - \gamma_j} [1 - (\gamma_i - \gamma_j)] + e^{-\gamma_j} [1 - (-\gamma_j)] \\ &+ \gamma_0 + \sum_{i=1, i \neq j}^d \gamma_i e^{\gamma_i - \gamma_j} - \gamma_j \left(\sum_{k=1, k \neq j}^d e^{\gamma_k - \gamma_j} + e^{-\gamma_j} \right) \\ &= \sum_{i \neq j, i \neq j}^d e^{\gamma_i - \gamma_j} + e^{-\gamma_j} + \gamma_0. \end{aligned} \tag{A18}$$

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

$$\gamma_0 e^{\gamma_j} + \sum_{i=1, i \neq j}^d e^{\gamma_i} = -1, \quad j = 1, 2, \dots, d. \tag{A19}$$

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

$$Y = \begin{pmatrix} \gamma_0 & 1 & \cdots & 1 \\ 1 & \gamma_0 & 1 \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & \gamma_0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} e^{\gamma_1} \\ e^{\gamma_2} \\ \vdots \\ e^{\gamma_d} \end{pmatrix}, \quad -\mathbf{1} = \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix}. \tag{A20}$$

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 $i, j = 1, 2, \dots, d$. Consequently, for $i \neq j$, we get $B_{ij} = 1$, and by summing Equation (A14) over $i \in S$,

$$\begin{aligned} 0 &= \|\mathbf{u}\| \sum_{i=1}^d C_{2,i} = \sum_{i=1}^d \sum_{j=1, j \neq i}^d B_{ij} x_j^* - \sum_{i=1}^d x_i^* \left(\sum_{k=1, k \neq i}^d B_{ki} + z_i \right) - \|\mathbf{u}\| \\ &= - \sum_{i=1}^d x_i^* z_i - \|\mathbf{u}\|, \end{aligned} \tag{A21}$$

which can also be expressed by $z^T \mathbf{x}^* = \|\mathbf{u}\|$. 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,

$$\mathbf{B}^* = \begin{pmatrix} -\lambda & 1 & \cdots & 1 \\ 1 & -\lambda & 1 \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & 1 & -\lambda \end{pmatrix} \tag{A22}$$

for $\lambda = d - 1 + 1/\mathbb{E}[\mathcal{T}]$. 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, i.e., $\mathbb{H}(X) = H_1 + H_2 + H_3$ with

$$\begin{aligned} H_1 &= - \sum_{i=1}^d \beta_i \log \beta_i, \\ H_2 &= \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} z_j (1 - \log z_j), \text{ and} \\ H_3 &= \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d B_{ij} (1 - \log B_{ij}). \end{aligned} \tag{A23}$$

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 ([13] Sect. 5.3)

$$\mathbb{P}(E = j) = \frac{z_j x_j^*}{\|\mathbf{u}\|} \tag{A24}$$

to rewrite the second term as

$$H_2 = \sum_{j=1}^d \mathbb{P}(E = j) (1 - \log z_j) = \sum_{j=1}^d \mathbb{H}(T_E | E = j) \mathbb{P}(E = j) = \mathbb{H}(T_E | E), \tag{A25}$$

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 $\mathbb{E}[\mathcal{T}]$, we get $z_j = 1/\mathbb{E}[\mathcal{T}]$ for all $j \in S$.

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

$$H_3 = \sum_{j=1}^d \frac{x_j^*}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d 1 \cdot (1 - \log 1) = (d - 1) \sum_{j=1}^d \mathbb{E}[O_j] = (d - 1) \mathbb{E}[\mathcal{T}] \tag{A26}$$

by Equation (33). 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

$$\begin{aligned} H_3(\varepsilon) &= \sum_{j=1}^d \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d 1 \cdot (1 - \log 1) \left(1 - \mathbb{1}_{\{i=k, j=l\}}\right) \\ &\quad + \frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} (1 + \varepsilon) [1 - \log(1 + \varepsilon)] \\ &= \sum_{j=1}^d \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d \left(1 - \mathbb{1}_{\{i=k, j=l\}}\right) + \frac{x_l^*(\varepsilon)}{\|\mathbf{u}\|} (1 - \delta) \end{aligned} \tag{A27}$$

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

$$\begin{aligned} H_3(\varepsilon) &= \left[\sum_{j=1}^d \frac{x_j^*(\varepsilon)}{\|\mathbf{u}\|} \sum_{i=1, i \neq j}^d 1 \right] - \delta \frac{x_1^*(\varepsilon)}{\|\mathbf{u}\|} = (d-1) \sum_{j=1}^d \mathbb{E}[O_j(\varepsilon)] - \delta \frac{x_1^*(\varepsilon)}{\|\mathbf{u}\|} \\ &= (d-1) \mathbb{E}[\mathcal{T}(\varepsilon)] - \delta \frac{x_1^*(\varepsilon)}{\|\mathbf{u}\|} = (d-1) \mathbb{E}[\mathcal{T}] - \delta \frac{x_1^*(\varepsilon)}{\|\mathbf{u}\|} \\ &< H_3. \end{aligned} \tag{A28}$$

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

Proposition A3. Consider the set \mathcal{M}_2 of compartmental systems in equilibrium given by Equation (7) with a predefined nonzero input vector \mathbf{u} and a predefined positive steady-state vector \mathbf{x}^* . The compartmental system $M_2^* = M(\mathbf{u}, B^*)$ with $B^* = (B_{ij})_{i,j \in S}$ given by

$$B_{ij} = \begin{cases} \sqrt{\frac{x_i^*}{x_j^*}}, & i \neq j, \\ -\sum_{k=1, k \neq j}^d \sqrt{\frac{x_k^*}{x_j^*}} - \frac{1}{\sqrt{x_j^*}}, & i = j, \end{cases} \tag{A29}$$

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 Equation (A15) and setting $\partial L/\partial B_{ij} = 0$ leads to

$$-\log B_{ij} + \gamma_i - \gamma_j = 0, \quad i \neq j. \tag{A30}$$

An interchange of the indices and summing the two equations gives

$$\log B_{ij} + \log B_{ji} = 0. \tag{A31}$$

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

$$-\log z_j - \gamma_j = 0, \quad j \in S, \tag{A32}$$

and in turn $z_j = (x_j^*)^{-1/2}$. The 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

$$\frac{\partial^2}{\partial B_{ij}^2} \mathbb{H}(X) = \frac{\partial}{\partial B_{ij}} \frac{-x_j^*}{\|\mathbf{u}\|} \log B_{ij} = -\frac{x_j^*}{\|\mathbf{u}\| B_{ij}} < 0 \tag{A33}$$

and

$$\frac{\partial^2}{\partial z_j^2} \mathbb{H}(X) = \frac{\partial}{\partial z_j} \frac{-x_j^*}{\|\mathbf{u}\|} \log z_j = -\frac{x_j^*}{\|\mathbf{u}\| z_j} < 0. \tag{A34}$$

\square

References

1. Burnham, K.P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*; Springer: New York, NY, USA; Berlin/Heidelberg, Germany, 2002.
2. Höge, M.; Wöhling, T.; Nowak, W. A Primer for Model Selection: The Decisive Role of Model Complexity. *Water Resour. Res.* **2018**, *54*, 1688–1715. [[CrossRef](#)]

3. Golan, A.; Harte, J. Information theory: A foundation for complexity science. *Proc. Natl. Acad. Sci. USA* **2022**, *119*, e2119089119. [[CrossRef](#)] [[PubMed](#)]
4. Jost, J. *Dynamical Systems: Examples of Complex Behaviour*; Springer: Berlin, Germany; New York, NY, USA, 2005.
5. Fan, J.; Meng, J.; Ludescher, J.; Chen, X.; Ashkenazy, Y.; Kurths, J.; Havlin, S.; Schellnhuber, H.J. Statistical physics approaches to the complex Earth system. *Phys. Rep.* **2021**, *896*, 1–84. [[CrossRef](#)] [[PubMed](#)]
6. Anderson, D.H. *Compartmental Modeling and Tracer Kinetics*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 1983; Volume 50.
7. Walter, G.G.; Contreras, M. *Compartmental Modeling with Networks*; Birkhäuser: Basel, Switzerland, 1999.
8. Haddad, W.M.; Chellaboina, V.; Hui, Q. *Nonnegative and Compartmental Dynamical Systems*; Princeton University Press: Princeton, NJ, USA, 2010.
9. Eriksson, E. Compartment Models and Reservoir Theory. *Annu. Rev. Ecol. Syst.* **1971**, *2*, 67–84. [[CrossRef](#)]
10. Bolin, B.; Rodhe, H. A note on the concepts of age distribution and transit time in natural reservoirs. *Tellus* **1973**, *25*, 58–62. [[CrossRef](#)]
11. Rasmussen, M.; Hastings, A.; Smith, M.J.; Agosto, F.B.; Chen-Charpentier, B.M.; Hoffman, F.M.; Jiang, J.; Todd-Brown, K.E.O.; Wang, Y.; Wang, Y.P.; et al. Transit times and mean ages for nonautonomous and autonomous compartmental systems. *J. Math. Biol.* **2016**, *73*, 1379–1398. [[CrossRef](#)]
12. Sierra, C.A.; Müller, M.; Metzler, H.; Manzoni, S.; Trumbore, S.E. The muddle of ages, turnover, transit, and residence times in the carbon cycle. *Glob. Chang. Biol.* **2017**, *23*, 1763–1773. [[CrossRef](#)]
13. Metzler, H.; Sierra, C.A. Linear Autonomous Compartmental Models as Continuous-Time Markov Chains: Transit-Time and Age Distributions. *Math. Geosci.* **2018**, *50*, 1–34. [[CrossRef](#)]
14. Metzler, H.; Müller, M.; Sierra, C.A. Transit-time and age distributions for nonlinear time-dependent compartmental systems. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 1150–1155. [[CrossRef](#)]
15. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
16. Jaynes, E.T. Information Theory and Statistical Mechanics. II. *Phys. Rev.* **1957**, *108*, 171–190. [[CrossRef](#)]
17. Pressé, S.; Ghosh, K.; Lee, J.; Dill, K.A. Principles of maximum entropy and maximum caliber in statistical physics. *Rev. Mod. Phys.* **2013**, *85*, 1115. [[CrossRef](#)]
18. Pesin, Y.B. Characteristic Lyapunov exponents and smooth ergodic theory. *Uspekhi Mat. Nauk* **1977**, *32*, 55–112.
19. Dehmer, M.; Mowshowitz, A. A history of graph entropy measures. *Inf. Sci.* **2011**, *181*, 57–78. [[CrossRef](#)]
20. Trucco, E. A note on the information content of graphs. *Bull. Math. Biol.* **1956**, *18*, 129–135. [[CrossRef](#)]
21. Morzy, M.; Kajdanowicz, T.; Kazienko, P. On Measuring the Complexity of Networks: Kolmogorov Complexity versus Entropy. *Complexity* **2017**, *2017*, 3250301. [[CrossRef](#)]
22. Bonchev, D.; Buck, G.A. Quantitative Measures of Network Complexity. In *Complexity in Chemistry, Biology, and Ecology*; Springer: Berlin/Heidelberg, Germany, 2005; pp. 191–235.
23. Shannon, C.E.; Weaver, W. *The Mathematical Theory of Communication*; The University of Illinois Press: Urbana, IL, USA, 1949.
24. Jaynes, E.T. Macroscopic prediction. In *Complex Systems—Operational Approaches in Neurobiology, Physics, and Computers*; Springer: Berlin/Heidelberg, Germany, 1985; pp. 254–269.
25. Roach, T.N.F. Use and Abuse of Entropy in Biology: A Case for Caliber. *Entropy* **2020**, *22*, 1335. [[CrossRef](#)] [[PubMed](#)]
26. Aoki, I. Entropy laws in ecological networks at steady state. *Ecol. Model.* **1988**, *42*, 289–303. [[CrossRef](#)]
27. Haddad, W.M. A Unification between Dynamical System Theory and Thermodynamics Involving an Energy, Mass, and Entropy State Space Formalism. *Entropy* **2013**, *15*, 1821–1846. [[CrossRef](#)]
28. Haddad, W.M. *A Dynamical Systems Theory of Thermodynamics*; Princeton University Press: Princeton, NJ, USA, 2019.
29. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.
30. Bad Dumitrescu, M.E. Some informational properties of Markov pure-jump processes. *Časopis Pro Pěstování Mat.* **1988**, *113*, 429–434.
31. Gaspard, P.; Wang, X.J. Noise, chaos, and (ϵ, τ) -entropy per unit time. *Phys. Rep.* **1993**, *235*, 291–343. [[CrossRef](#)]
32. Jacquez, J.A.; Simon, C.P. Qualitative theory of compartmental systems. *Siam Rev.* **1993**, *35*, 43–79. [[CrossRef](#)]
33. Norris, J.R. *Markov Chains*; Cambridge University Press: Cambridge, UK, 1997.
34. Neuts, M.F. *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*; The Johns Hopkins University Press: Baltimore, MD, USA, 1981.
35. Albert, A. Estimating the Infinitesimal Generator of a Continuous Time, Finite State Markov Process. *Ann. Math. Stat.* **1962**, *33*, 727–753. [[CrossRef](#)]
36. Doob, J.L. *Stochastic Processes*; Wiley: New York, NY, USA, 1953; Volume 7.
37. Walter, G.G. Size identifiability of compartmental models. *Math. Biosci.* **1986**, *81*, 165–176. [[CrossRef](#)]
38. Bellman, R.; Åström, K.J. On structural identifiability. *Math. Biosci.* **1970**, *7*, 329–339. [[CrossRef](#)]

39. Emanuel, W.R.; Killough, G.G.; Olson, J.S. Modelling the Circulation of Carbon in the World's Terrestrial Ecosystems. In *Carbon Cycle Modelling*; SCOPE 16; John Wiley and Sons: Hoboken, NJ, USA, 1981; pp. 335–353.
40. Sierra, C.A.; Quetin, G.R.; Metzler, H.; Müller, M. A decrease in the age of respired carbon from the terrestrial biosphere and increase in the asymmetry of its distribution. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2023**. [[CrossRef](#)]
41. Thompson, M.V.; Randerson, J.T. Impulse response functions of terrestrial carbon cycle models: Method and application. *Glob. Chang. Biol.* **1999**, *5*, 371–394. [[CrossRef](#)]
42. Wang, Y.P.; Chen, B.C.; Wieder, W.R.; Leite, M.; Medlyn, B.E.; Rasmussen, M.; Smith, M.J.; Agosto, F.B.; Hoffman, F.; Luo, Y.Q. Oscillatory behavior of two nonlinear microbial models of soil carbon decomposition. *Biogeosciences* **2014**, *11*, 1817–1831. [[CrossRef](#)]
43. Ågren, G.I. Investigating soil carbon diversity by combining the MAXimum ENTropy principle with the Q model. *Biogeochemistry* **2021**, *153*, 85–94. [[CrossRef](#)]
44. 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. Mathematical Modeling to Estimate Photosynthesis: A State of the Art. *Appl. Sci.* **2022**, *12*, 5537. [[CrossRef](#)]
45. Manzoni, S.; Porporato, A. Soil carbon and nitrogen mineralization: Theory and models across scales. *Soil Biol. Biochem.* **2009**, *41*, 1355–1379. [[CrossRef](#)]
46. 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. Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate Free-Air CO₂ Enrichment studies. *New Phytol.* **2014**, *202*, 803–822. [[CrossRef](#)]
47. Farquhar, G.D.; von Caemmerer, S.v.; Berry, J.A. A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta* **1980**, *149*, 78–90. [[CrossRef](#)]
48. Friedlingstein, P.; Cox, P.; Betts, R.; Bopp, L.; von Bloh, W.; Brovkin, V.; Cadule, P.; Doney, S.; Eby, M.; Fung, I.; et al. Climate–carbon cycle feedback analysis: Results from the C4MIP model intercomparison. *J. Clim.* **2006**, *19*, 3337–3353. [[CrossRef](#)]
49. Friedlingstein, P.; Meinshausen, M.; Arora, V.K.; Jones, C.D.; Anav, A.; Liddicoat, S.K.; Knutti, R. Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *J. Clim.* **2014**, *27*, 511–526. [[CrossRef](#)]
50. Todd-Brown, K.E.; Randerson, J.T.; Post, W.M.; Hoffman, F.M.; Tarnocai, C.; Schuur, E.A.; Allison, S.D. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **2013**, *10*, 1717–1736. [[CrossRef](#)]
51. Ceballos-Núñez, V.; Müller, M.; Sierra, C.A. Towards better representations of carbon allocation in vegetation: A conceptual framework and mathematical tool. *Theor. Ecol.* **2020**, *13*, 317–332. [[CrossRef](#)]
52. Sierra, C.A.; Harmon, M.E.; Perakis, S.S. Decomposition of heterogeneous organic matter and its long-term stabilization in soils. *Ecol. Monogr.* **2011**, *81*, 619–634. [[CrossRef](#)]
53. Ge, H.; Pressé, S.; Ghosh, K.; Dill, K.A. Markov processes follow from the principle of maximum caliber. *J. Chem. Phys.* **2012**, *136*, 064108. [[CrossRef](#)]
54. Girardin, V. Entropy Maximization for Markov and Semi-Markov Processes. *Methodol. Comput. Appl. Probab.* **2004**, *6*, 109–127. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.