Information content and maximum entropy of compartmental systems in equilibrium

The authors introduce new indices linked to information theory, with application to continuous-time Markov chains in the field of compartmental systems. The examples and applications appear as a mere illustration of the theoretical results.

This manuscript is conceptualized as a first step to introduce entropy measures derived from information theory to deterministic compartmental systems via a stochastic interpretation of the systems. We lay down the mathematical foundations of those new entropy measures and display their main properties and potential future applications at the hand of several rather simple examples. As also mentioned below, in the new version of the manuscript, we added two paragraphs as to how the understanding of entropy in simple models can help understand more complex systems. Actual applications to more complex systems can and will be considered in future manuscripts, once the mathematical foundations are settled and published.

Compartmental models are not new in Markov chains theory, visited compartments are just visited states, as some of the references point out.

Still, in information theory, as explained in the annex, non absorbing chains are considered, with a likelihood with respect to a classical Borel measure, and the entropy rate is lim\_T H(X\_t, 0<t<T)/T. This quantity is well-known to inherit the properties of Shannon entropy for random variables.

Here, the Markov chains are absorbing ones, so the usual notion of entropy rate par unit time does not apply. Instead, the authors consider a reference measure induced by all the possible trajectories of the chain before its absorption. They determine the likelihood of the model with respect to this measure, and then compute the Shannon entropy, called H(X).

The new indices are:

H(X)/E T, where T is the time to absorption; this is not a rate per unit time (which Shannon entropy rate is).

H(X)/E N, where N is the number of jumps before absorption.

We thank the reviewer for the concise summary of one mathematical novelty in our manuscript, in particular the consideration of Shannon entropy of absorbing Markov chains as opposed to classical theory on non-absorbing Markov chains.

The properties of Shannon entropy for random variables (convexity, symmetry, etc.) are the basis of the MaxEnt method. These properties are compulsory, allowing for a unique maximum, uniformity, and identifiability. Any new index that is to be used for MaxEnt has to be proven to satisfy these properties. Unfortunately, this would be very difficult to obtain when dividing entropy by

the expectation of a quantity depending on the process itself.

We agree that those properties are crucial for our newly introduced entropy quantities in order to be useful in the general context of MaxEnt. We also agree that from the definitions made in the manuscript such properties could not simply be inferred. Consequently, we added Proposition 2 to Section 3.1 (old version: Section 2.5) which proves our entropy rate per jump of an absorbing Markov chain to equal the classical entropy rate of a stationary process that describes an indefinite journey of one particle through the system and immediate jump back into the system after leaving it. The entropy rate per unit time follows then just as a renormalization associated to the average time between jumps. This shows that our entropy rates introduced for absorbing Markov chains are indeed classical entropy rates of suitable stationary stochastic processes. Hence, they possess all the desirable properties an entropy rate should possess and can readily be used in the framework of MaxEnt.

The examples could bring some arguments for using these new indices. Unfortunately, nothing helps the reader to understand whether their values confirm some known facts on the behavior of the systems, or if the comments are just comments of the figures in the tables and graphs.

First, the examples stand for themselves and serve as means for a better understanding of the new entropy concept itself. Second, they are supposed to allow the reader to extrapolate this local understanding to other systems.

We added two paragraphs to the “Discussion” section in which we explain how a good understanding of the entropy of two otherwise well-studied carbon cycle models adds to our understanding of more complex systems. By virtue of our new entropy measures it becomes rather easy to understand why the among-model agreement for carbon uptake via photosynthesis is much higher than the among-model agreement for soil organic carbon cycling. Both the slower cycling speed and the higher heterogeneity of processes in soils contribute to an inherently higher system uncertainty against which all predictive models must fight. We also mention now that a higher global surface temperature, following Example 4.2 (old version: 3.2) is likely to decrease the future predictability of the global carbon cycle.

Section 3.1 is a clear illustration of the complexity of the maximum: a theoretical higher maximum for a higher number of states tells nothing on the real level of entropy of a given system.

This is exactly what we wanted to convey in this section.

Furthermore, line 579 *Usually, entropy is maximized when the system is highly symmetric*, this is not usual by chance but as a direct consequence of the structural definition of entropy. The same is true on the comments in line 592 actually linked to the structural properties of the function x log x.

We removed the word “Usually” and pointed out the underlying reasons for the mentioned properties as remarked by the reviewer.

No mathematical argument is given in the paper to ensure that the following claim (line 336) holds, not even some empirical explanation is given that could justify a conjecture.

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

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

The interpretation of the path entropy as uncertainty of the entire path follows directly from its definition. We thank the reviewer to point out that this is not the case for the two entropy rates. As mentioned in an earlier comment, the new Proposition 2 in Section 3.1 of the revised manuscript shows that our entropy rates are just classical entropy rates of suitable stationary processes (per jump: discrete time, per unit time: continuous time), which corroborates our statement by the very definition of these classical quantities.

The discussion and conclusion following the examples show that indeed these conditions are not satisfied. Commenting on the poor results for Example 3.4 that

*This example is only supposed to give a first impression of how the maximum entropy principle can be used in combination with entropy rates or path entropy in similar situations. Practical examples usually have a high level of complexity such that existence and uniqueness of a maximum entropy model have to be studied on a case-by-case basis.*

is just the opposite of the powerful concept of entropy that is known to fit any situation when handled pertinently, and actually the opposite of any scientific tool.

Indeed, we cannot guarantee that a local maximum entropy rate is also a global one. The reason for that is not the lack of required properties of our introduced entropy rates (as mentioned earlier), but that the parameter space is not guaranteed to be convex. The properties of the parameter space depend on the measurements and experiment at hand and vary independently of the mathematical theory. Optimization over non-convex spaces is a complex mathematical problem in general and finding optimal solutions is far beyond the scope of this manuscript. We provide the new entropy quantities equipped with all desirable properties such that standard optimization methods can be applied to them. We do not solve the general problems of local versus global optima and existence and uniqueness of global maxima in optimization over non-convex spaces.

We thank the reviewer to make us aware that the approach of running one local optimization to find a global maximum in the original manuscript was too simple-minded. In the new version we place a grid over a relevant-assumed section of the parameter space and run more than 450,000 local optimizations starting from each grid point. We identify our global maximum candidate as the maximum of all those local maximizations. Even though mathematical theory does not allow us to conclude that our global maximum candidate is indeed a (or even ‘the’) global maximum, the new Figure 6, replacing the old one showing only one local optimization path, gives clear indications that our global maximum candidate is a good and reasonable choice. Furthermore, this example now demonstrates one possible (brute force) way of dealing with optimization over a non-convex parameter set and can thus serve as a first role model for practical cases in higher dimensions arising from real-world experimental measurements. We adapted the corresponding part in the discussions section accordingly.

The level of mathematics and probability necessary for a reader to understand the paper is rather high, especially on Markov theory and information theory. Therefore, many lengthy sections on the very basics of entropy is of no use, and indeed most of the material is used nowhere in the paper. Since the authors are no experts in information theory, a reference to Cover and Thomas, and some

definitions closely linked to the paper would be more pertinent than Sections 2.1 and A1, beginning of Section 2.6, and so on.

We removed the entire appendix A1 on basic information theory. Instead, we refer the reader to the associated chapters in Cover and Thomas and to Bad-Dumitrescu in the introduction of Section 2. There we kept only the notions and properties of entropy that are fundamental to the further understanding of the manuscript, even though they might seem too basic at first glance. We believe that the average reader is (like us) not an expert in information theory and will appreciate a short introduction to the topic that emphasizes the links to the new entropy concepts that we introduce later in the text.

Of course, Kolmogorov, topological or graphical entropies (lines 59-71) have nothing to do here, Markov processes are linked to Shannon entropy from his original paper giving birth to information theory. By the way, the determination of the reference measure is much less classical and could be detailed.

We agree that Markov processes are immediately linked to Shannon entropy. However, compartmental systems are classically considered deterministic dynamical systems. Hence, Kolmogorov, topological, and graphical entropies are the classical entropy measures that come to mind. We explain why those classical approaches fail for compartmental systems, and it is only by linking compartmental systems to Markov chains through considering the stochastic travel of a single particle that Shannon entropy becomes the natural choice of uncertainty measure. This is a major novelty of the manuscript. We made this point clearer now in the introduction.

The derivation of the reference measure is presented in all its detail Section 2.5. We do not know how to make it any more detailed. The main idea is to introduce a reference measure for paths of finite length and then extend this measure to infinite paths with measure-theoretic technicalities that do not add to the understanding of the simple idea for finite paths. We included it already in the original manuscript mainly for the sake of completeness.

Some local comments:

--The end of Section 1 is repeated at the beginning of Section 2. Section 2 is far too long, mixing definitions, main theoretical results and examples.

We rewrote the end of Section 1 to avoid repetition with the beginning of Section 2. We furthermore split Section 2 into two new sections (Section 2 and Section 3). The first one, Section 2, now contains the bare minimum of mathematical background necessary to understand the new entropy measures. These new measures are then introduced in the second new Section 3 alongside their links to MaxEnt and structural model identification.

--Strangely, right in the middle of the paper, Section 2.7 is devoted to the estimation of the number of compartments, with no link to entropy, even though its title is *Structural model identification via MaxEnt*.

We changed the section title to “Structural model identification assisted by MaxEnt”.

Only the first paragraph in Section 2.7 is devoted to the estimation of number of compartments. The number is a crucial first step for further model identification and cannot be assisted by MaxEnt, as explained in the paragraph. One possible tool that can help identify the necessary number of pools is the transfer function of the system. This exact function is also crucial to further MaxEnt-assisted identification of the model structure, such as external inputs and outputs and links between different compartments. We tried to make the connection clearer in the manuscript by adding a connecting sentence.

-- Why does MaxCal only appear in the conclusion? It would have been welcome in the introduction, as a method linked to both entropy and Markov chains.

We discern our theory now already in the introduction from MaxCal to avoid initial confusion around the term “path entropy”.

All in all, my advice is to reject the paper, not finalized enough for publication.

We are very grateful to the reviewer for his/her thorough reading of the manuscript and his/her insightful remarks. Not only did he/she appreciate the novelty of our approach, he/she also provided us with very useful comments which allowed us to significantly improve the manuscript in its display of the fundamental properties of the newly introduced entropy measures, in their application to the problem of structural model selection, and in their interpretation on a scale larger than the presented examples.