Response to reviewer comments

Comment 1.1 *In their review, the authors undertake a challenge when they combine meteorological observations and a mathematical definition in terms of advanced statistical and dynamical approaches. While a general definition of persistence is already hard, the related notion recurrence is even more difficult to grasp and to distinguish. Not surprisingly, the review is weak when the authors attempt the almost impossible task to provide a comprehensive theoretical framework, but it gets stronger when it resorts to methodologies and meteorological examples. In particular the variety of examples and their properties show the width of the seemingly simple notion persistence.*
The authors are well aware of the difficulties involved and do not hide that. The review is worth to read for two reasons: the collection of methodologies that have been suggested to analyze persistence, and the huge amount of examples. In summary the authors provide a broad and useful overview with a lot of insight. Below I mention some aspects that I noticed.

Answer: Thank you for your positive comments. It has indeed not been an easy task to collect and arrange the various methods and definitions that have been used to describe the extremely broad concept of persistence. The goal of sections 2 and 3 is to introduce, with as few mathematical formalities and assumptions as possible, the different meanings that have been given to persistence, and the questions one should ask themselves when using this word. We realise that this discussion has its limitations, and a bit more formalism (as you suggested in your Comment 1.6 below, for instance, would probably be welcome.

Comment 1.2 *Line 55: Since persistence needs a timescale for the definition this question is somehow circular: "At which timescale(s) does the persistence occur?" There could be reference to Section 3.4.*

Answer: You are right that the current formulation sounds a bit circular, so we could reformulate such as "What are the persistent timescales in the data?" As you say, persistence requires a timescale to be defined, but this timescale is not necessarily known or define a priori.

Comment 1.3 *Line 124, Eq* $(\mathbf{x}(t))_t \in R^m$ *what is the meaning of the subscript t together with the argument t? This should be explained.*

Answer: Thanks for pointing this out. We are simply referring to the dynamical system here, and wanted to point out that it was indexed by time, but this notation is unnecessarily complicated. We will replace by $\mathbf{x}(t)$.

Comment 1.4 *Line 126, I recommend to use 'state space' instead of phase space (the often used notion phase space is reserved for Hamiltonian systems).*

Answer: This is a very good suggestion which we will adopt in the revised version. Thanks!

Comment 1.5 *Line 136: To characterize the difference between precipitation and temperature an autocorrelation time would be appropriate here. This is better than 'inertia' since it is* *difficult to associate precipitation with an inertia, while temperature, on the other hand, has something like an inertia due to the heat capacity.*

Answer: This example is meant as an illustration of the concept of global persistence (before introducing it a few lines later). Autocorrelation would indeed be very relevant to characterise the difference between temperature and precipitation (and it is one of the common "global persistence" methods). The term "inertia" is used here in a more general way than in the context of the temperature/precipitation example, to highlight the fact that global persistence of a dynamical system describes how "fast" it tends to evolve, in other words it refers to its inertia.

Comment 1.6 *Lines 143, 155 in Section 3.1 Global, state and episodic persistence. It seems that global and state persistence are different concepts. Global persistence appears nothing else than stationarity which includes recurrence. Why is global persistence 'strongly related to intrinsic system predictability'? Maybe these definitions could be useful: Does persistence of a state x(t) mean dx(t)/dt=0. And could global persistence be defined in terms of integrals or averages like* $d < x(t) > /dt = 0$*? And are conservation laws useful?*

Answer: Global and state persistence are indeed two different concepts. Global persistence relates to the stationarity/recurrence of a time series, but assessed over all the time steps of the series, while state persistence relates to the stationarity/recurrence of a specific system state. Your suggestion to illustrate these definitions with equations is very useful! In that sense, global persistence of a system $\mathbf{x}(t)$ looks at whether $\left\langle \frac{d\mathbf{x}(t)}{dt} \right\rangle = 0$ is, while the (state) persistence of system state $\mathbf{x}_0(t)$ looks at whether $\left\langle \frac{d\mathbf{x}(t)}{dt}\Big|_{\mathbf{x}(t)=\mathbf{x}_0} \right\rangle = 0$ (with $<>$ being a time average). Whether conservation laws are useful in general is hard to say. It is quite possible that persistent system states or persistent periods are characterised by the (quasi)-conservation of some global quantity, but the answer really depends on the system under analysis.

Comment 1.7 *Line 184: What is a 'symmetrically, persistent state'?*

Answer: This is a misunderstanding, "symmetrically" does not refer to the state but is used to highlight the connection to the previous sentence. To avoid this problem, we can replace "symmetrically" by "correspondingly".

Comment 1.8 *Line 188: How can this sentence be understood: 'However, state persistence only characterizes the average behavior of system states.'*

Answer: What we mean by this sentence is that state persistence relates to the overall tendency of a given system state to be persistent, and does not imply that all occurrences of this state will necessarily be persistent. For instance, extreme warm conditions in London may on average last a week, which can be considered to be persistent, and in that case the state "very warm conditions" would be a persistent state of the time series of daily temperatures in London, but it may happen that extreme warm conditions last only one or two days. We will reformulate the manuscript to make this clearer.

Comment 1.9 *Line 201: Section 3.2 Here is a good opportunity to define Lagrangian stationarity of a quantity* $\psi(\mathbf{x}, t)$ *along a flow* $\mathbf{u}$*, by* $\partial\psi/\partial t + \mathbf{u} \cdot \nabla\psi = 0$*, in comparison to the Eulerian stationarity with* $\partial\psi/\partial t = 0$.

Answer: Thanks for this excellent suggestion, which we will make sure to adopt in the revision.

Comment 1.10 *Line 216: The authors write 'self-similarity of system values* $\mathbf{x}(t)$ *with a metric' but this notion can be confused with geometric self-similarity used in the definition of fractal objects for example. Or is that what the authors intend?*

Answer: We hadn't thought of that, but you are right, it could be confusing for some readers. What we mean by "self-similarity" is the tendency for successive values of a dynamical system to be similar to each other, similar being understood in the sense of "resemblance". It is quite difficult to find an alternative. The use of "similarity" is well-accepted in e.g. "similarity metrics" which quantify the similarity, or likeness, between two objects. However, we could add a note to make it clear that our use of the word is different from its acceptation in geometry.

Comment 1.11 *Line 300: write what the symbol E(. . .) denotes.*

Answer: $\mathcal{E}$ denotes the expectancy with respect to the distribution of X_t. We will make sure to specify it in the revision.

Comment 1.12 *Line 348: On Long-range memory: mention the absence of a timescale.*

Answer: We mention this on l.355 when discussing the limitations of this approach.

Comment 1.13 *Line 360: Write that the Hurst exponent H and d are related by H=2d+1, see the Table 1 in Franzke et al. (2020).*

Answer: Thanks for the suggestion, we will adopt it in the revision.

Comment 1.14 *Line 382: Written is: 'If temporal dependence is present'. This is unclear. Eq. (10) is the result a power-law in of* ρ *(Eq.(6)), line 350.*

Answer: We are not sure what you mean by this comment. What we wanted to say is that for pure random noise, the spectral density is constant, while for series with non-zero autocorrelation (at least at small lags), the power spectrum will often (but not necessarily) exhibit a power-law dependence as a function of frequency.

Comment 1.15 *Line 490: Section 4.2.4. The role of the extremal index* θ *is difficult here. In extreme value statistics the extremal index is the inverse time scale with co-occurring extreme events, used to eliminate short term variability and to combine events. Why is this index a measure of stationarity? Is it correct that short term excursions are allowed with* θ *in Eq (13)?*

Answer: You are correct that in extreme value statistics, θ is typically used to decluster extreme event occurrences in peak-over-threshold analysis. However, the use of the extremal index as a measure of temporal persistence has been introduced by Lucarini et al. (2016) and has been extensively used by the same authors in a large number of papers. The idea is that a high θ indicates that event occurrences tend to be isolated, while a low θ indicates that events occur as part of a larger group. Equation (13) relates to a marginal probability on $\mathbf{x}(t)$ – it does not explicitly describe the length of event clusters. However, in extreme value theory it can be shown that θ defined as in Equation (13) is approximately equal to the inverse of the mean cluster size (e.g., Moloney et al. `https://doi.org/10.1063/1.5079656`. Short-term excursions are allowed in the sense that the relationship between θ and the mean cluster size is approximate.

Comment 1.16 *Line 569: Please write that R_{X_0} is the residence time for state x_0.*

Answer: Thanks for the suggestion, we will adopt it in the revision.