Correspondence between temporal correlations in time series, inverse problems, and the spherical model

Riccardo Marcaccioli[1,*] and Giacomo Livan [1,2,†]

[1]*Department of Computer Science, University College London, 66-72 Gower Street, London WC1E 6EA, United Kingdom*
[2]*Systemic Risk Centre, London School of Economics and Political Sciences, Houghton Street, London WC2A 2AE, United Kingdom*

In this paper we employ methods from statistical mechanics to model temporal correlations in time series. We put forward a methodology based on the maximum entropy principle to generate ensembles of time series constrained to preserve part of the temporal structure of an empirical time series of interest. We show that a constraint on the lag-one autocorrelation can be fully handled analytically and corresponds to the well-known spherical model of a ferromagnet. We then extend such a model to include constraints on more complex temporal correlations by means of perturbation theory, showing that this leads to substantial improvements in capturing the lag-one autocorrelation in the variance. We apply our approach on synthetic data and illustrate how it can be used to formulate expectations on the future values of a data-generating process.

I. INTRODUCTION

During last two decades, multidisciplinary applications of physics—ranging from economics [1,2] and finance [3,4] to sociology [5,6], biology [7,8], and linguistics [6,9]—have witnessed an increasing attention from the physics community [10–13]. Indeed, physicists have contributed to the development of methodologies that are of crucial importance to such disciplines, such as, e.g., network modeling and analysis [14,15], game theory [16,17], and time series analysis [18,19].

Arguably, one of the main driving forces of such an interest is statistical mechanics, which provides a unified and coherent framework based on first principles to model large interacting systems even outside the realm of physics. In particular, as originally suggested by Jaynes [20,21], the maximum entropy principle has been used as a flexible tool to build unbiased statistical models in a vast range of different disciplines [22–24].

However, in most of such applications, the maximum entropy principle is used in the opposite way with respect to its classical use in statistical mechanics, where the goal is usually to compute observable macroscopic quantities (such as correlations in an Ising model) from the unobservable microscopic laws ruling the interactions between the components of a system [25]. The opposite problem is that of inferring the parameters of an interacting system (e.g., the coupling constants and fields in an Ising model) from snapshots of its microscopic configurations. This is referred to as the "inverse problem." In physics it has received considerable attention, especially when applied to fully connected Ising models [25,26]. Outside physics, instead, it has provided a theoretical basis for some of the aforementioned interdisciplinary applications, due to the

increased accessibility of the "microscopic configurations" of many nonphysical systems (e.g., financial markets, social networks, neuron firing patterns, etc.).

Jaynes proposed an alternative principle when dealing with models of time-evolving systems, typically nonstationary or out-of-equilibrium ones. Known as the maximum caliber principle [27], its goal is to determine an unbiased distribution over all possible paths of a system by maximizing the system's path entropy while preserving some desired constraints on its trajectories. Researchers have used the maximum caliber principle in a wide rage of different applications [28,29], the majority of which have been devoted to determining the transition rates of Markov models [30,31] for systems evolving in continuous time between a fixed set of states.

In the case of systems evolving in discrete time, the maximum caliber principle can be shown to coincide with the maximum entropy principle by mapping time as the spatial dimension of a lattice whose sites are occupied by events [32,33]. In practice, this mapping effectively corresponds to *time series*, as even most systems evolving in continuous time are sampled at discrete times.

In Ref. [33], we have shown how the entropy formulation can be used to generate ensembles of multivariate time series in discrete time constrained to preserve, on average, some empirically observed distributional properties of a multivariate system (such as higher order moments and seasonalities). One of the main challenges of the multivariate case presented in Ref. [33] is that of explicitly accounting for correlations, which can be captured only indirectly via other constraints.

In the present work, we partially overcome such limitations by tackling the problem of explicitly accounting for temporal correlations in the case of univariate systems. Modeling the temporal correlations of statistical systems is a notoriously challenging task. The most frequently used tools are autoregressive models belonging to the ARCH-GARCH

family [34,35] or stochastic processes such as the Ornstein-Uhlenbeck model [36]. These models—and those inspired by them—have enjoyed great success in a variety of applications where modeling time correlations can be crucial, such as in economics or finance [37,38]. Here, instead, we adopt a data-driven perspective—grounded in the maximum entropy principle—in order to capture the time correlations of a system without the need to explicitly model its time evolution.

First, we will briefly introduce the general methodology in its full mathematical form (Sec. II). Then we will apply it to generate ensembles of time series designed to preserve on average correlations between first moments as measured in an empirical time series of interest. In order to do so, we will leverage the spherical model [39,40], and we will show that it naturally corresponds to autoregressive processes [34] (Sec. III). After that, we will proceed to account for higher order temporal correlations. We will do so by solving a more complex model by expanding on the spherical model by means of perturbation theory (Sec. IV). In the former case, the analytical knowledge of the spherical model's partition function makes the calibration of the proposed approach extremely simple, whereas in the latter case we will show how the Plefka expansion [41]—a technique commonly used for the inverse Ising problem—can be applied to find an approximate solution.

II. MAXIMUM ENTROPY FRAMEWORK FOR TIME SERIES DATA

Let $\mathcal{X}$ be the set of all real-valued time series of length T, and let $\overline{X} \in \mathcal{X}$ be an empirical time series of interest, i.e., $\bar{x}_t$ stores the time t value sampled from a variable under consideration. The goal of the methodology is to define an ensemble able to preserve, as ensemble averages, L empirical measurements on $\overline{X}$. In other words, we want to find a probability density function $P(X)$ over $\mathcal{X}$, such that the expectation values $\langle \mathcal{O}_\ell(X) \rangle = \sum_{X \in \mathcal{X}} \mathcal{O}_\ell(X) P(X)$ of a set of observables $(\ell = 1, \ldots, L)$ coincide with their values measured in the given time series $\overline{O}_\ell = \mathcal{O}_\ell(\overline{X})$. In these terms, the problem is ill-defined, as $P(X)$ may be defined in an arbitrary number of ways. However, if we require $P(X)$ to also maximize the entropy $S(X) = \sum_{X \in \mathcal{X}} -P(X) \ln(X)$, computing $P(X)$ becomes a constrained maximization problem which can be uniquely solved by choosing

$$P(X) = \frac{e^{-H(X)}}{Z},$$

where $H(X) = \sum_\ell \beta_\ell \, \mathcal{O}_\ell(X)$ is the Hamiltonian of the ensemble, $\beta_\ell \ (\ell = 1, \ldots, L)$ are Lagrange multipliers introduced to enforce the constraints, and $Z = \sum_X e^{-H(X)}$ is the partition function of the ensemble, which verifies $\langle \mathcal{O}_\ell(X) \rangle = -\partial \ln Z / \partial \beta_\ell \, , \forall \ell$. The existence and uniqueness of the Lagrange multipliers can be proved, and it can also be shown that they are equivalent to those that maximize the likelihood of drawing the time series $\overline{X}$ from the ensemble [42].

The problem of determining $P(X)$ has therefore been solved. However, explicitly computing the Lagrange multipliers β_ℓ that maximize the likelihood of drawing the data from the ensemble without an analytical form for Z can be achieved only by means of Boltzmann learning gradient-descent al-

gorithms [25]. These ultimately require an exhaustive phase space exploration through sequential Monte Carlo simulations, which quickly becomes computationally unfeasible for $T \gg 1$. Therefore, finding a closed-form solution (even an approximate one) for Z is the cardinal problem to be solved in order to fully define a working methodology.

As a dummy example to illustrate how a specific ensemble can be computed, let us consider an empirical time series $\overline{X}_t$ of length T and let us choose as constraints its sample mean $\overline{m} = \sum_{t=1}^{T} \bar{x}_t / T$ and mean-square value $\overline{V} = \sum_{t=1}^{T} \bar{x}_t^2 / T$. In order to compute the partition function Z, let us denote as x_t the tth element in X, and let us place each of such elements on a one dimensional lattice of length T. The constraints on the mean and mean-square value lead to the following Hamiltonian:

$$H = \sum_{t=1}^{T} \left[\lambda_1 x_t + \lambda_2 x_t^2 \right].$$

After having specified the constraints, what is left to do is to evaluate the partition function. In order to do that, we need to properly define the sum over the phase space $\mathcal{X}$ appearing in the definition of Z:

$$Z = \sum_{X \in \mathcal{X}} e^{-H(X)} = \int_{-\infty}^{+\infty} \prod_{t=1}^{T} dx_t \, e^{-H(X)}$$

$$= \prod_{t=1}^{T} \int_{-\infty}^{+\infty} dx_t \, e^{-\lambda_1 x_t - \lambda_2 x_t^2} = \left(\sqrt{\frac{\pi}{\lambda_2}} e^{\frac{\lambda_1^2}{4\lambda_2}} \right)^T, \ \lambda_2 > 0.$$

Once the partition function is known, the Lagrange multipliers λ_1 and λ_2 can be found by solving the following system of coupled equations:

$$\overline{m} = -\frac{1}{T} \frac{\partial \ln Z}{\partial \lambda_1} = -\frac{\lambda_1}{2\lambda_2}$$

$$\overline{V} = -\frac{1}{T} \frac{\partial \ln Z}{\partial \lambda_2} = \frac{\lambda_1^2 + 2\lambda_2}{4\lambda_2^2},$$

which leads to the following probability density function for the ensemble:

$$P(X) = \left[\frac{1}{2\pi(\overline{V} - \overline{m}^2)} \right]^{T/2} \prod_{t=1}^{T} e^{-\frac{(x_t - \overline{m})^2}{2(\overline{V} - \overline{m}^2)}}, \quad V > m^2,$$

which is the factorized probability density function of T independent Gaussian random variables with mean $\overline{m}$ and variance $(\overline{V} - \overline{m})^2$.

III. THE FIRST HAMILTONIAN

We shall now apply the framework introduced in the previous section to a more complex set of constraints, namely, the sample mean $(\overline{m})$, mean-square value $(\overline{V})$, and temporal correlation at lag one $\overline{C}_1 = \sum_{t=1}^{T} \bar{x}_t \bar{x}_{t+1}$ (notice that the following steps generalize to a generic temporal correlation $\overline{C}_\tau = \sum_{t=1}^{T} \bar{x}_t \bar{x}_{t+\tau}$).

Let us place the data points on a one-dimensional temporal lattice, whose sites $t = 1, \ldots, T$ correspond to the events of a time series of interest $\bar{x}_1, \ldots, \bar{x}_T$. After doing that, the

specified set of constraints leads to the following Hamiltonian:

$$H = \sum_{t=1}^{T} \left[\lambda_1 x_t + \lambda_2 x_t^2 + \lambda_3 x_t x_{t+1} \right], \qquad (1)$$

where we are assuming spherical boundary conditions $x_{T+1} = x_1$. The Hamiltonian in Eq. (1) is that of the spherical model [39], a well-known model in statistical mechanics.

Having specified the Hamiltonian, the task now becomes finding the partition function Z, which reads

$$Z = \int_{-\infty}^{+\infty} \prod_{t=1}^{T} dx_t \ e^{-\lambda_1 x_t - \lambda_2 x_t^2 - \lambda_3 x_t x_{t+1}}$$

$$= \int d^T x \ e^{-x^{\mathrm{T}} A x + B^{\mathrm{T}} x} = \sqrt{\frac{\pi^T}{\det A}} \ e^{\frac{B^{\mathrm{T}} A^{-1} B}{4}}, \qquad (2)$$

where we have introduced the following vector notation:

$$B^{\mathrm{T}} = -\lambda_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$$A = \begin{pmatrix} \lambda_2 & \lambda_3 & 0 & \cdots & \cdots & \cdots & \cdots & \lambda_3 \\ \lambda_3 & \lambda_2 & \lambda_3 & 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \lambda_3 & \lambda_2 & \lambda_3 & \ddots & & & \vdots \\ \vdots & 0 & \ddots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 & \vdots \\ \vdots & & & \ddots & \lambda_3 & \lambda_2 & \lambda_3 & 0 \\ \vdots & & & & 0 & \lambda_3 & \lambda_2 & \lambda_3 \\ \lambda_3 & 0 & \cdots & \cdots & \cdots & 0 & \lambda_3 & \lambda_2 \end{pmatrix}.$$

Using the fact that A is a special case of a real symmetric circulant matrix (whose spectral properties are generally known [43]), one can show that its eigenvalues are $\Lambda_t = \lambda_2 + \lambda_3 \cos \frac{2\pi}{T}(t-1)$ $(t = 1, \ldots, T)$. These can be used to highlight, in the limit $T \gg 1$, the explicit dependency of Z from the Lagrange multipliers. Expanding each term appearing in Eq. (2), we have

$$\det A = \prod_{t=1}^{T} \Lambda_t = e^{\sum_{t=1}^{T} \ln \left[\lambda_2 + \lambda_3 \cos \frac{2\pi}{T}(t-1) \right]}$$

$$\approx e^{\frac{T}{2\pi} \int_0^{2\pi} d\omega \ \ln \left[\lambda_2 + \lambda_3 \cos \omega \right]} = e^{T \ln \frac{\lambda_2 + \sqrt{\lambda_2^2 - \lambda_3^2}}{2}}$$

$$= \left(\frac{\lambda_2 + \sqrt{\lambda_2^2 - \lambda_3^2}}{2} \right)^T,$$

$$B^{\mathrm{T}} A^{-1} B = \lambda_1^2 \ b^{\mathrm{T}} A^{-1} b = \lambda_1^2 \ b^{\mathrm{T}} \frac{1}{\Lambda_1} b = T \ \frac{\lambda_1^2}{\lambda_2 + \lambda_2}, \qquad (3)$$

where we have used the fact that $b = (1, \ldots, 1)$ is the eigenvector of A (and therefore of A^{-1}) associated to Λ_1. Plugging the above expressions into Eq. (2), we obtain the ensemble's partition function, which reads

$$Z = \left(\frac{2\pi}{\lambda_2 + \sqrt{\lambda_2^2 - \lambda_3^2}} \right)^{\frac{T}{2}} e^{T \frac{\lambda_1^2}{4(\lambda_2 + \lambda_2)}}. \qquad (4)$$

From Eq. (4) we can derive the system of equations for the Lagrange multipliers:

$$\frac{\overline{m}}{T} = -\frac{\lambda_1}{2(\lambda_2 + \lambda_3)},$$

$$\frac{\overline{V}}{T} = \frac{\lambda_1^2}{4(\lambda_2 + \lambda_3)^2} + \frac{1}{2\sqrt{\lambda_2^2 - \lambda_3^2}},$$

$$\frac{\overline{C_1}}{T} = \frac{\lambda_1^2}{4(\lambda_2 + \lambda_3)^2} + \frac{\lambda_3}{2(\lambda_3^2 - \lambda_2^2 + \lambda_2 \sqrt{\lambda_2^2 - \lambda_3^2})}. \qquad (5)$$

The above equations can be easily solved analytically. Their expressions are not particularly instructive, so we omit them for easiness of exposition. Once the system in Eq. (5) has been solved the ensemble is fully defined, and instances can be drawn from it with standard Monte Carlo methods [44].

Figure 1 shows an application of the ensemble aimed at reconstructing a known data-generating process. Black lines correspond to data generated synthetically from an autoregressive model defined as follows: $Y_{t+1} = \xi_t^{(a_1, a_2)} Y_t + \xi_t^{(a_3, a_4)}$, where $\xi_t^{(a,b)}$ is a random number drawn at time t from a uniform distribution in the interval $[a, b]$. The solid black line corresponds to the final 30 points of an initial time series of length $T = 180$, which we use to compute the Lagrange multipliers appearing in Eq. (4). The black dashed line corresponds to the continuation of such time series beyond time T, which we use both to update the Lagrange multipliers in "real time," and to test the agreement between the scenarios generated by the ensemble with respect to new data points. The blue (light gray) solid line and shaded region correspond to "out of sample" next-step expectations for times $t > T$ (i.e., obtained by recomputing the ensemble's Lagrange multipliers for all times $t \geqslant T + 1$), denoting, respectively, the average value and 99% confidence interval computed from the ensemble via Monte Carlo simulations. The purple (dark grzy) solid line and shaded region instead capture the "true" next-step evolution of the system. They correspond, respectively, to the mean and 99% confidence interval computed over a sample of 10^6 trajectories of the aforementioned autoregressive model generated as one-step increments starting, at all times, from the values represented by the dashed black line. As can be seen from a qualitative inspection of Fig. 1, the ensemble reproduces rather faithfully the average time evolution of the underlying data-generating process. There are, however, some visible deviations between the two confidence intervals shown in Fig. 1. These are due to the fact that the data-generating process has non trivial time correlations in its higher order moments, which are not captured by the ensemble. These will be captured by the model introduced in the next section, where we will also perform a more rigorous statistical assessment of the model's ability to reconstruct a data-generating process.

IV. A MORE COMPLEX HAMILTONIAN

We now proceed to investigate a more complex ensemble encoding additional constraints. We consider the following
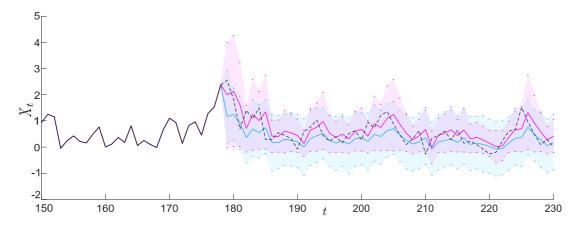
FIG. 1. Reconstruction of a known data-generating process. Black lines correspond to data generated synthetically from the autoregressive model $Y_{t+1} = \xi_t^{(0,1.5)} Y_t + \xi_t^{(-0.3,0.7)}$. The solid black line corresponds to data up to time $T = 180$, which are used to compute the initial values of the ensemble's Lagrange multipliers appearing in Eq. (4), and the black dashed line corresponds to the evolution of the process beyond time T. The blue (light gray) solid line and shaded region denote, respectively, "out of sample" next-step expectations for times $t > T$ based on the ensemble, with Lagrange multipliers updated in "real time" based on new data points. The purple (dark gray) solid line and shaded region correspond, respectively, to the mean and 99% confidence interval computed over a sample of 10^6 trajectories of the process X_t generated as one-step increments starting, at all times, from the values represented by the dashed black line.

Hamiltonian:

$$H = \sum_{t=1}^{T} \left[\lambda_1 x_t + \lambda_2 x_t^2 + \lambda_3 x_i x_{t+1} + \lambda_4 x_t^2 x_{t+1}^2 + \lambda_5 x_t^4 \right], \quad (6)$$

which enforces the constraints already considered in the Hamiltonian of Eq. (1), plus additional constraints on the sample mean fourth power ($\sum_{t=1}^{T} \bar{x}_t^4$) and on the time correlations at lag one between squared values ($\sum_{t=1}^{T} \bar{x}_t^2 \bar{x}_{t+1}^2$). Such constraints—coupled with the ones mentioned previously—effectively amount to constraining, respectively, the ensemble average on the kurtosis and on the variance autocorrelation at lag one.

Similarly to Eq. (2), the partition function resulting from Eq. (6) reads

$$Z = \int_{-\infty}^{+\infty} \prod_{t=1}^{T} dx_i \, e^{-\lambda_1 x_t - \lambda_2 x_t^2 - \lambda_3 x_t x_{t+1} + \lambda_4 x_t^2 x_{t+1}^2 + \lambda_5 x_t^4}. \quad (7)$$

Integrals similar to the one above appear in $\lambda \phi^4$ lattice field theories and are known for not being solvable analytically. However, such calculations are commonly tackled by using resummation techniques or perturbation theory [45]. Following this line of research, we will make use of the Plefka expansion—a perturbation method widely used in the inverse Ising problem—in order to find approximate estimates of the true Lagrange multipliers.

In standard perturbation theory, the true Hamiltonian H of a system is written as a sum of an unperturbed part H_0 and a perturbation H_p: $H = H_0 + H_p$. Using this notation, the partition function of the system becomes

$$Z = \sum_{\mathcal{X}} e^{-(H_0 + H_p)} = Z_0 \sum_{\mathcal{X}} \frac{e^{-H_0}}{Z_0} e^{-H_p}$$

$$= Z_0 \langle e^{-H_p} \rangle_0 = Z_0 \sum_k \frac{(-1)^k}{k!} \langle H_p^k \rangle_0, \quad (8)$$

where Z_0 is the partition function of the unperturbed system ($Z_0 = \sum_{\mathcal{X}} e^{-H_0}$), and $\langle \cdots \rangle_0$ is the average over the ensemble defined by Z_0. Equation (8) is exact. However, in order to make it usable in practice, one needs to truncate the power series expansion (which becomes a power series expansion in the Lagrange multipliers appearing in the definition of H_p) to a certain order k. Of course, if one is lucky enough to find a recursion for $\langle H_p^k \rangle_0$ and to sum the resulting series, one can in principle compute the true partition function Z.

The Plefka expansion follows a very similar procedure to the one just described. It starts from the Hamiltonian of the system written as $H = H_0 + \lambda H_p$, where λ is a constant that serves to distinguish different perturbation orders which will be ultimately set to one. Instead of expanding the partition function Z, the Plefka expansion considers the free energy of the system:

$$F = -\ln Z = -\ln Z_0 - \ln \frac{Z}{Z_0} = F_0 + F_p, \quad (9)$$

where F_0 is the free energy of the unperturbed ensemble and $F_p = -\ln \frac{Z}{Z_0}$. We can now expand F_p as a power series in λ:

$$F_p = -\lambda f_1 + \frac{\lambda^2}{2} f_2 - \frac{\lambda^3}{3!} f_3 + \cdots, \quad (10)$$

where we used the fact that if $\lambda = 0$ then $F = F_0$. Substituting into $e^{-F_p} = Z/Z_0$, we obtain

$$\frac{Z}{Z_0} = 1 - \lambda f_1 + \frac{\lambda^2}{2} (f_2 + f_1^2) - \frac{\lambda^3}{3!} (f_3 + f_1^3 + 3 f_2 f_1) + \cdots. \quad (11)$$

Comparing Eq. (11) with the direct power series expansion $Z/Z_0 = \sum_k (-\lambda)^k \langle H_p^k \rangle_0 / k!$, we obtain an explicit expression for every term of the expansion in Eq. (10):

$$\begin{aligned} f_1 &= \langle H_p \rangle_0, \\ f_2 &= \langle H_p^2 \rangle_0 - f_1^2, \\ f_3 &= \langle H_p^3 \rangle_0 - f_1^3 - 3 f_1 f_2. \end{aligned} \quad (12)$$

As can seen from Eq. (12), the expansion of the free energy F is effectively an expansion around the cumulants of the unperturbed ensemble. A similar idea was developed (eight years earlier than Plefka) by Bogolyubov *et al.* [46] for the ferromagnetic Ising model.

Let us now perform a second-order Plefka expansion in the case of the Hamiltonian in Eq. (6). We will consider the spherical model (4) as the unperturbed ensemble Z_0, with the perturbation given by $H_p = \sum_t [\lambda_4 x_t^2 x_{t+1}^2 + \lambda_5 x_t^2]$. As a result, the second-order approximated free energy reads

$$
\begin{aligned}
F \approx F_0 &- \sum_t \left[\lambda_5 \langle x_t^4 \rangle_0 + \lambda_4 \langle x_t^2 x_{t+1}^2 \rangle_0 \right] \\
&+ \frac{1}{2} \sum_{t,t'} \left[\lambda_5^2 \langle x_t^4 x_{t'}^4 \rangle_0 + \lambda_4^2 \langle x_t^2 x_{t+1}^2 x_{t'}^2 x_{t'+1}^2 \rangle_0 \right. \\
&\left. + 2\lambda_5 \lambda_4 \langle x_t^4 x_{t'}^2 x_{t'+1}^2 \rangle_0 \right] \\
&- \frac{1}{2} \sum_t \left[\lambda_5 \langle x_t^4 \rangle_0 + \lambda_4 \langle x_t^2 x_{t+1}^2 \rangle_0 \right]^2,
\end{aligned}
\tag{13}
$$

where the expansion above has introduced a second time index t'. In the following, we shall make use of this in order to introduce distances between sites $t - t'$, which correspond to temporal distances between events in the original time series.

We now proceed to evaluate the expectation values in Eq. (13) around the ensemble defined by Eq. (4). In order to do that, we need to apply Isserlis' theorem [47], a result which is also largely employed in quantum field theory under the name Wick's theorem [48].

For easiness of exposition, let us redefine some quantities appearing in Eq (5) as follows:

$$
\begin{aligned}
m &= -\frac{\lambda_1}{2(\lambda_2 + \lambda_3)} = -\frac{\partial}{\partial \lambda_1} \ln Z_0, \\
s_0 &= \frac{1}{2\sqrt{\lambda_2^2 - \lambda_3^2}} = -\frac{\partial}{\partial \lambda_2} \ln Z_0 \bigg|_{\lambda_1 = 0}, \\
s_1 &= \frac{\lambda_3}{2\left(\lambda_3^2 - \lambda_2^2 + \lambda_2 \sqrt{\lambda_2^2 - \lambda_3^2}\right)} = -\frac{\partial}{\partial \lambda_3} \ln Z_0 \bigg|_{\lambda_1 = 0}, \\
s_{tt'} &= \langle x_t x_{t'} \rangle_0 |_{\lambda_1 = 0},
\end{aligned}
\tag{14}
$$

where $s_0 = s_{tt}$ and $s_1 = s_{t,t+1}$, $\forall t$.

We can now proceed to calculate the expectation values appearing in Eq. (13). These read

$$
\begin{aligned}
\langle x_t^4 \rangle_0 &= m^4 + 6m^2 s_0 + 3s_0^2, \\
\langle x_t^2 x_{t+1}^2 \rangle_0 &= (m^2 + s_0)^2 + 4m^2 s_1 + 2s_1^2, \\
\langle x_t^4 x_{t'}^4 \rangle_0 &= (m^4 + 6m^2 s_0 + 3s_0^2)^2 + 16(m^3 + 3ms_0)^2 s_{tt'} \\
&\quad + 72(m^2 + s_0)^2 s_{tt'}^2 + 96m^2 s_{tt'}^3 + 24 s_{tt'}^4, \\
\langle x_t^4 x_{t'}^2 x_{t'+1}^2 \rangle_0 &= f(m, s_0, s_1, s_{tt'}, s_{t,t'+1}), \\
\langle x_t^2 x_{t+1}^2 x_{t'}^2 x_{t'+1}^2 \rangle_0 &= g(m, s_0, s_1, s_{tt'}, s_{t,t'+1}, s_{t+1,t'}),
\end{aligned}
\tag{15}
$$

where f and g are polynomial functions of their variables and are specified in the Appendix.

As one can see from Eqs. (13) and (15), the second-order approximation contains the covariances of the unperturbed

Hamiltonian at all possible ranges, i.e., not just at lag one. As a result, we need to find an explicit form for $s_{tt'}$ in order to move forward. Following the steps that lead to the solution of the Gaussian integral in Eq. (2), we have

$$
\begin{aligned}
s_{tt'} &= \langle x_t x_{t'} \rangle_0 |_{\lambda_1 = 0} = \left\langle \sum_s V_{ts} y_s \sum_k V_{t'k} y_k \right\rangle_0 \bigg|_{\lambda_1 = 0} \\
&= \sum_{s,k} V_{ts} V_{t'k} \langle y_s y_k \rangle_0 |_{\lambda_1 = 0} = \sum_s V_{ts} V_{t's} \langle y_s^2 \rangle_0 |_{\lambda_1 = 0} \\
&= \sum_s \frac{1}{2T} \frac{\cos\left[\frac{2\pi}{T}(s-1)(t-t')\right]}{\lambda_2 + \lambda_3 \cos\left[\frac{2\pi}{T}(s-1)\right]},
\end{aligned}
\tag{16}
$$

where $V_{tk} = \frac{1}{\sqrt{T}}(\cos\left[\frac{2\pi}{T}(t-1)(k-1)\right] + \sin\left[\frac{2\pi}{T}(t-1)(k-1)\right])$ is the tth element of the kth eigenvector of the matrix A in Eq. (2), and $y_k = \sum_t V_{tk} x_t$. The above expression can be then rewritten as

$$
\begin{aligned}
s_{tt'} = \sum_s \frac{1}{2T} \cos\left[\frac{2\pi}{T}(s-1)R\right] \\
\times \int_0^\infty dz\, e^{-z\{\lambda_2 + \lambda_3 \cos\left[\frac{2\pi}{T}(s-1)\right]\}},
\end{aligned}
\tag{17}
$$

where $R = t - t'$ is the distance between the two lattice sites being considered. We can now approximate the above expression for $T \gg 1$ as follows:

$$
\begin{aligned}
s_{tt'} &\overset{T \gg 1}{\approx} \int_0^\infty \frac{dz}{2} e^{-z\lambda_2} \int_0^{2\pi} \frac{d\omega}{2\pi} e^{-z\lambda_3 \cos \omega} \cos(\omega R) \\
&= \int_0^\infty \frac{dz}{2} e^{-z\lambda_2} I_R(-\lambda_3 z) \\
&= \frac{\left(-\frac{\lambda_3}{|\lambda_3|}\right)^R}{2} \int_0^\infty dz\, e^{-z\lambda_2} I_R(|\lambda_3| z) \\
&\overset{\lambda_2 - |\lambda_3| \ll 1}{\approx} \frac{[-\text{sign}(\lambda_3)]^R}{2} \int_0^\infty dz\, \frac{e^{-z\lambda_2 + |\lambda_3| z - \frac{R^2}{2|\lambda_3| z}}}{\sqrt{2\pi |\lambda_3| z}} \\
&= \frac{[-\text{sign}(\lambda_3)]^R}{2|\lambda_3|\sqrt{2\frac{\lambda_2}{|\lambda_3|} - 2}} e^{-|R|\sqrt{2\frac{\lambda_2}{|\lambda_3|} - 2}},
\end{aligned}
\tag{18}
$$

where $I_n(x)$ is the modified Bessel function of the first kind. Let us briefly comment on the approximation made in the third step of the above expression (i.e., for $\lambda_2 - |\lambda_3| \ll 1$). As it can be seen from the expressions for s_0 and s_1 in Eq. (14), such an approximation corresponds to a regime of strong time correlations up to lag one. The approximation effectively becomes useful only to compute time correlations at lag two or higher, i.e., to compute $s_{tt'}$ for $t' > t + 1$, given that those at lower lags are known exactly. Therefore, $\lambda_2 - |\lambda_3| \approx 1$ corresponds to a regime of low time correlations even at lags one and zero (it should be noted here that correlations of the type $\langle x_t x_{t'} \rangle$ are not normalized to one when $t = t'$, as is instead the case with the standard definition of autocorrelation). This, in turn, ensures that time correlations at higher lags will be low enough to make the error due to the above approximation negligible.

We can now plug the above result into Eq. (13) via Eq. (15) in order to compute the approximate form of the free energy
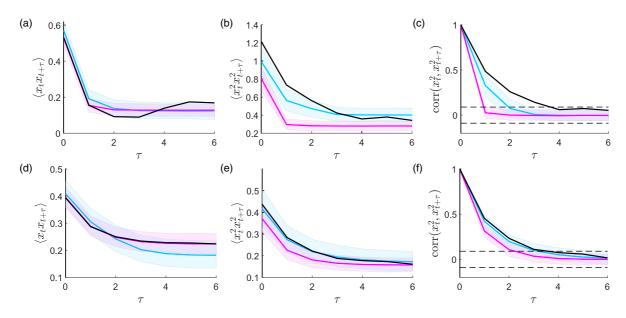
FIG. 2. Ability of the different models to match specified constraints. Comparisons between empirical time correlations and the corresponding quantities as measured in the ensembles defined by Eq. (4) [purple (dark gray)] and Eq. (7) [blue (light gray)]. Panels (a) and (d) refer to $\langle x_t x_{t+\tau} \rangle$, panels (b) and (e) to $\langle x_t^2 x_{t+\tau}^2 \rangle$, and panels (c) and (f) to $(\langle x_t^2 x_{t+\tau}^2 \rangle - \langle x_t^2 \rangle \langle x_{t+\tau}^2 \rangle)/\langle x_t^4 \rangle$. In the three upper panels the empirical correlations (black solid lines) are computed from one instance of the autoregressive model $Y_{t+1} = \xi_t^{(-1.5,1.5)} Y_t + \xi_t^{(-0.2,0.8)}$, whereas in the three lower panels correlations are computed from the model $Y_{t+1} = \xi_t^{(-0.375,1.125)} Y_t + \xi_t^{(-0.2,0.8)}$. In panels (c) and (f) horizontal dashed lines denote the 95% confidence level interval for the autocorrelation of white noise.

deriving from the partition function in Eq. (7). After having computed such approximate form for F, we can calculate the Lagrange multipliers as usual, i.e., by solving the system of equations $\langle \mathcal{O}_\ell(X) \rangle = \partial F / \partial \beta_\ell$, $\forall \ell$. Alternatively, one could truncate Eq. (11) to the second order of the couplings λ_3 and λ_4, find an approximate form Z_p of Z, and then maximize the approximate likelihood $e^{-H(\bar{X})}/Z_p$.

In Fig. 2 we show the ability of the ensemble introduced in this section to match the imposed constraints with respect to its unperturbed counterpart. We do so using two autoregressive models with markedly distinct correlation features. The first model ($Y_{t+1} = \xi_t^{(-1.5,1.5)} Y_t + \xi_t^{(-0.2,0.8)}$ [Figs. 2(a) and 2(c)] is designed to produce time series that, on average, have nonzero correlations only between second or higher order moments. This represents an "adversarial" example, in the sense that the only correlations present in the process cannot be captured by the unperturbed model of Eq. (4). This would suggest the need to use a "stronger" perturbation than the second-order one in order to substantially improve the model's ability to capture the correlations of the process. However, Figs. 2(b) and 2(c) show that even stopping the perturbation expansion at the second order gives a sizable improvement. The second model ($Y_{t+1} = \xi_t^{(-0.375,1.125)} Y_t + \xi_t^{(-0.2,0.8)}$ [2(d)–2(f)] is instead designed to produce time series with time correlations between first moments as well. This represents a scenario where the unperturbed model captures by design correlations between first moments, which translates into a partial ability to capture higher order correlations. These are then fully captured by the full model [see Figs. 2(e) and 2(f)].

In order to quantitatively assess the improvement of the proposed perturbative solution with respect to the unperturbed ensemble, we report the results from a few simple predic-

tion exercises in Table I. Namely, we seek to predict the mean and the 10% and 90% quantiles of the data-generating process one lag ahead. We compare such quantities against those computed from both the unperturbed and full ensemble, and we quantify the agreement with two widely adopted metrics of accuracy: the root-mean-square error (RMSE) and the R^2. As can be seen, in all cases switching from the unperturbed ensemble to the one in Eq. (7) systematically provides a measurable improvement, regardless of the specific model considered. Notably, the biggest relative improvement occurs for the model that we identified as an "adversarial" example. This is because, as mentioned

TABLE I. Accuracy of one lag ahead predictions of the mean and 10% and 90% quantiles of the three data-generating processes used so far. These are denoted, respectively, as M_1 ($Y_{t+1} = \xi_t^{(-0.375,1.125)} Y_t + \xi_t^{(-0.2,0.8)}$), M_2 ($Y_{t+1} = \xi_t^{(-1.5,1.5)} Y_t + \xi_t^{(-0.2,0.8)}$), and M_3 ($Y_{t+1} = \xi_t^{(0,1.5)} Y_t + \xi_t^{(-0.3,0.7)}$). The means and quantiles are denoted as $\bar{\mathbf{x}}$ and $\bar{\mathbf{q}}$. H_1 and H_2 denote, respectively, predictions obtained by means of the unperturbed ensemble of Eq. (4) and the full ensemble of Eq. (7).

	M_1		M_2		M_3	
	RMSE	R^2	RMSE	R^2	RMSE	R^2
$\bar{x}_{H_1}$	0.0267	0.995	0.125	0.923	0.204	0.938
$\bar{x}_{H_2}$	0.0155	0.998	0.0818	0.969	0.176	0.954
$q_{H_1}^{0.9}$	0.0900	0.943	0.236	0.847	0.277	0.866
$q_{H_2}^{0.9}$	0.0511	0.985	0.122	0.957	0.232	0.910
$q_{H_1}^{0.1}$	0.0825	0.960	0.206	0.887	0.170	0.970
$q_{H_2}^{0.1}$	0.0495	0.975	0.188	0.906	0.151	0.976

above, in that case the unperturbed model cannot capture by design the relevant time correlations in the data-generating process.

V. CONCLUSIONS

In this paper we have shown how we can apply tools from classical statistical mechanics to time series analysis by simply mapping the time dimension of a time-evolving system onto the spatial dimension of a lattice. This allows us to design ensembles that preserve, on average, some desired constraints on the temporal structure of the data under consideration by imposing constraints on such a lattice.

In particular, we have shown how a constraint on the lag-one autocorrelation corresponds to a well-known ensemble in statistical mechanics, namely, the spherical model of a (anti-) ferromagnet. Moreover, we have shown how the spherical model can be used as a basis to handle higher order temporal correlations by using a perturbation theory approach. We have also shown how the inferred Lagrange multipliers of the ensembles can be updated in "real time" as new data

from the system of interest are collected, which in turn allows us to obtain information about the possible evolution of the underlying data-generating process.

The framework presented here, coupled with tools commonly used to tackle inverse Ising problems such as pseudolikelihood methods, can be adapted in order to handle multiple time series and their correlations. Moreover, the accuracy of the single time series case presented here can be improved by considering higher perturbation orders or by considering different Hamiltonians. In particular, it would be interesting to extend the approach proposed here to Hamiltonians whose Lagrange multipliers are drawn from parametric distributions (similarly to couplings in spin-glass systems), which would provide an alternative—and possibly even more flexible—method to fit ensembles to some desired constraints. We hope to see some of these topics pursued in the near future.

APPENDIX: POLYNOMIAL FUNCTIONS USED IN EQ. (15)

In the following we report in full the expressions of the functions f and g introduced in Eq. (15)

$$
\begin{aligned}
\langle x_t^4 x_{t'}^2 x_{t'+1}^2 \rangle_0 = {} & m^8 + 12m^2 s_0^3 + 2m^4 \big[4(m^2 + 2s_1)s_{tt'} + s_1(2m^2 + s_1) + 6s_{tt'}^2 \big] \\
& + 8m^2 s_{(t+1)t'} \big[m^4 + 6(m^2 + s_1)s_{tt'} + 2m^2 s_1 + 6s_{tt'}^2 \big] + 12s_{(t+1)t'}^2 \big[m^4 + 2s_{tt'}(2m^2 + s_{tt'}) \big] \\
& + 4s_0 \big\{ 2m^2(m^2 + s_{tt'} + s_{(t+1)t'})(m^2 + 3s_{tt'} + 3s_{(t+1)t'}) + 3m^2 s_1^2 + 6s_1 [m^4 + 2m^2 s_{tt'} + 2s_{(t+1)t'}(m^2 + s_{tt'})] \big\} \\
& + 2s_0^2 \big[8m^4 + 6(2m^2 s_{tt'} + 2m^2 s_{(t+1)t'} + s_{tt'}^2 + s_{(t+1)t'}^2) + 6m^2 s_1 + 3s_1^2 \big] + 3s_0^4 \\
\langle x_t^2 x_{t+1}^2 x_{t'}^2 x_{t'+1}^2 \rangle_0 = {} & m^8 + 4s_0^3 m^2 + 16s_1^3 m^2 + 8s_1 [(m^2 + 2s_{(t+1)t'})(m^2 + 2s_{t(t'+1)}) \\
& + (2m^2 + s_{(t+1)t'} + s_{t(t'+1)})s_{(t+1)(t'+1)} + s_{tt'}(2m^2 + s_{(t+1)t'} + s_{t(t'+1)} + 4s_{(t+1)(t'+1)})]m^2 \\
& + s_0^4 + 4s_1^4 + 2s_0^2 \big[3m^4 + 4s_1 m^2 + 2s_{tt'} m^2 + 2s_{(t+1)t'} m^2 + 2s_{t(t'+1)} m^2 + 2s_{(t+1)(t'+1)} m^2 \\
& + 2s_1^2 + s_{tt'}^2 + s_{(t+1)t'}^2 + s_{t(t'+1)}^2 + s_{(t+1)(t'+1)}^2 \big] + 4s_1^2 [5m^4 + 4(s_{t(t'+1)} + s_{(t+1)(t'+1)})m^2 \\
& + 4s_{(t+1)t'}(m^2 + s_{t(t'+1)}) + 4s_{tt'}(m^2 + s_{(t+1)(t'+1)})] \\
& + 2 \big\{ [s_{(t+1)(t'+1)}^2 + 2(m^2 + 2s_{t(t'+1)})s_{(t+1)(t'+1)} + s_{t(t'+1)}(2m^2 + s_{t(t'+1)})]m^4 \\
& + 2s_{(t+1)t'}[m^4 + 2s_{t(t'+1)}(2m^2 + s_{t(t'+1)}] + 2(m^2 + 2s_{t(t'+1)})s_{(t+1)(t'+1)} \big\}m^2 \\
& + s_{(t+1)t'}^2 [m^4 + 2s_{t(t'+1)}(2m^2 + s_{t(t'+1)})] + 2s_{tt'}(2s_{(t+1)(t'+1)}^2 m^2 \\
& + \{ m^2 + 2s_{(t+1)t'})(m^2 + 2s_{t(t'+1)})m^2 + 4[m^2 + s_{(t+1)t'})(m^2 + s_{t(t'+1)})s_{(t+1)(t'+1)}] \\
& + s_{tt'}^2 (m^4 + 2s_{(t+1)(t'+1)}(2m^2 + s_{(t+1)(t'+1)})] \big\} + 4s_0 [2s_1^2 m^2 \\
& + [s_{tt'}^2 + 2(m^2 + s_{(t+1)t'} + s_{t(t'+1)})s_{tt'} + s_{(t+1)t'}^2 \\
& + (m^2 + s_{t(t'+1)} + s_{(t+1)(t'+1)})^2 + 2s_{(t+1)t'}(m^2 + s_{(t+1)(t'+1)})]m^2 + 2s_1 \{ 2(m^2 + s_{(t+1)t'} + s_{t(t'+1)})m^2 \\
& + s_{tt'}[2m^2 + s_{(t+1)t'} + s_{t(t'+1)}) + (2m^2 + s_{(t+1)t'} + s_{t(t'+1)})s_{(t+1)(t'+1)}] \big\}.
\end{aligned}
$$

[1] D. Challet, M. Marsili, and R. Zecchina, Phys. Rev. Lett. **84**, 1824 (2000).

[2] M. Bardoscia, G. Livan, and M. Marsili, J. Stat. Mech.: Theory Exp. (2017) 043401.

[3] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing: From Statistical Physics to Risk Management* (Cambridge University Press, Cambridge, 2003).

[4] R. N. Mantegna and H. E. Stanley, *Introduction to Econophysics: Correlations and Complexity in Finance* (Cambridge University Press, Cambridge, 1999).

[5] D. B. Bahr and E. Passerini, J. Math. Sociol. **23**, 1 (1998).

[6] C. Castellano, S. Fortunato, and V. Loreto, Rev. Mod. Phys. **81**, 591 (2009).

[7] G. Sella and A. E. Hirsh, Proc. Natl. Acad. Sci. USA **102**, 9541 (2005).

[8] H. J. Jensen, *Self-organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*, Cambridge lecture notes in physics Vol. 10 (Cambridge University Press, Cambridge, 1998).

[9] M. Cassandro, P. Collet, J. A. Galves, and C. Galves, Physica A **263**, 427 (1999).

[10] R. K. Pan, S. Sinha, K. Kaski, and J. Saramäki, Sci. Rep. **2**, 551 (2012).

[11] M. Szell, Y. Ma, and R. Sinatra, Nat. Phys. **14**, 1075 (2018).

[12] F. Battiston, F. Musciotto, D. Wang, A.-L. Barabási, M. Szell, and R. Sinatra, Nat. Rev. Phys. **1**, 89 (2019).

[13] R. Sinatra, P. Deville, M. Szell, D. Wang, and A.-L. Barabási, Nat. Phys. **11**, 791 (2015).

[14] M. Newman, *Networks* (Oxford University Press, Oxford, 2018).

[15] R. Marcaccioli and G. Livan, Nat. Commun. **10**, 745 (2019).

[16] Y. Sato, E. Akiyama, and J. D. Farmer, Proc. Natl. Acad. Sci. USA **99**, 4748 (2002).

[17] A. Melbinger, J. Cremer, and E. Frey, Phys. Rev. Lett. **105**, 178101 (2010).

[18] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).

[19] R. Morales, T. Di Matteo, R. Gramatica, and T. Aste, Physica A **391**, 3180 (2012).

[20] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[21] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).

[22] G. Cimini, T. Squartini, F. Saracco, D. Garlaschelli, A. Gabrielli, and G. Caldarelli, Nat. Rev. Phys. **1**, 58 (2019).

[23] S. J. Phillips, R. P. Anderson, and R. E. Schapire, Ecol. Model. **190**, 231 (2006).

[24] T. Watanabe, S. Hirose, H. Wada, Y. Imai, T. Machida, I. Shirouzu, S. Konishi, Y. Miyashita, and N. Masuda, Nat. Commun. **4**, 1370 (2013).

[25] H. C. Nguyen, R. Zecchina, and J. Berg, Adv. Phys. **66**, 197 (2017).

[26] Y. Roudi, J. Tyrcha, and J. Hertz, Phys. Rev. E **79**, 051915 (2009).

[27] E. T. Jaynes, Annu. Rev. Phys. Chem. **31**, 579 (1980).

[28] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Rev. Mod. Phys. **85**, 1115 (2013).

[29] P. D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. **148**, 010901 (2018).

[30] H. Ge, S. Pressé, K. Ghosh, and K. A. Dill, J. Chem. Phys. **136**, 064108 (2012).

[31] G. Stock, K. Ghosh, and K. A. Dill, J. Chem. Phys. **128**, 194102 (2008).

[32] S. Marzen, D. Wu, M. Inamdar, and R. Phillips, arXiv:1008.2726 (2010).

[33] R. Marcaccioli and G. Livan, Scientific Reports **10**, 10656 (2020).

[34] J. D. Hamilton, *Time Series Analysis* (Princeton: Princeton University Press, 1994).

[35] R. Engle, J. Econ. Perspect. **15**, 157 (2001).

[36] G. E. Uhlenbeck and L. S. Ornstein, Phys. Rev. **36**, 823 (1930).

[37] P. Embrechts, R. Frey, and H. Furrer, by D. Shanbag, and C. Rao **19**, 365 (2001).

[38] R. F. Engle, S. M. Focardi, and F. J. Fabozzi, *Encyclopedia of Financial Models* (Wiley Online Library, 2012).

[39] T. H. Berlin and M. Kac, Phys. Rev. **86**, 821 (1952).

[40] M. N. Barber and M. E. Fisher, Ann. Phys. **77**, 1 (1973).

[41] T. Plefka, J. Phys. A: Math. Gen. **15**, 1971 (1982).

[42] D. Garlaschelli and M. I. Loffredo, Phys. Rev. E **78**, 015101(R) (2008).

[43] R. M. Gray *et al.*, Found Trends Commun. Inf. Theory **2**, 155 (2006).

[44] K. Binder, D. Heermann, L. Roelofs, A. J. Mallinckrodt, and S. McKay, Comput. Phys. **7**, 156 (1993).

[45] G. Münster, Scholarpedia **5**, 8613 (2010).

[46] N. M. Bogolyubov, V. F. Brattsev, A. N. Vasil'ev, A. Korzhenevskii, and R. Radzhabov, Teor. Matemat. Fizika **26**, 341 (1976).

[47] L. Isserlis, Biometrika **12**, 134 (1918).

[48] G. C. Wick, Phys. Rev. **80**, 268 (1950).