Journal of Computer Applications in Archaeology

The Past as a Stochastic Process

POSITION PAPER

DAVID H. WOLPERT
MICHAEL H. PRICE
STEFANI A. CRABTREE
TIMOTHY A. KOHLER
JÜRGEN JOST
JAMES EVANS
PETER F. STADLER
HAJIME SHIMAO
MANFRED D. LAUBICHLER

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Historical processes manifest remarkable diversity. Nevertheless, scholars have long attempted, with some success, to identify patterns and categorize historical actors and influences. A stochastic process framework provides a structured approach for the analysis of large historical datasets that allows for detection of sometimes surprising patterns, identification of relevant causal actors both endogenous and exogenous to the process, and comparison between different historical cases. The combination of data, analytical tools and the organizing theoretical framework of stochastic processes complements traditional narrative approaches in history and archaeology.

CORRESPONDING AUTHOR:
David H. Wolpert
Santa Fe Institute, US
david.h.wolpert@gmail.com

KEYWORDS:
Historical trajectories; Social evolution; Computational history; stochastic processes; time series data sets

TO CITE THIS ARTICLE:
Wolpert, DH, Price, MH, Crabtree, SA, Kohler, TA, Jost, J, Evans, J, Stadler, PF, Shimao, H and Laubichler, MD. 2024. The Past as a Stochastic Process. *Journal of Computer Applications in Archaeology,* 7(1): 134–152. DOI: https://doi.org/10.5334/jcaa.113

What determines how history unfolds? Are the dynamics of history determined by the agency of individual actors? Or, in contrast, are they determined by broad-scale processes that individuals must react to but cannot affect? As a third possibility, are the dynamics all contingent, a succession of interacting events? Taking a more synthetic perspective, is there some way to unify all these possibilities in a formal framework, one that can account for a combination of factors jointly governing the dynamics of human societies?

Social scientists have not found it easy to find a consensus on such fundamental issues. At least as far back as Ibn Khaldun, observers of history hypothesized that history followed simple, universal dynamic rules (Alatas 2013). Similarly, in the 19th century, partly building on the work of Kant, thinkers like Hegel, Comte, Marx, Morgan, and Tylor emphasized necessity, generality, natural laws, and impersonal trends (McAllister 2002). They saw history as progressing in a law-like manner through well-defined stages, to approach some final destination. In the 20th century, while still ascribing a prominent role to impersonal trends, authors like Spengler speculated about the inevitable decline of societies, removing the assumption of "inherent progress" made in those 19th-century investigations. Both of these perspectives view impersonal processes — what some contemporary social scientists refer to as "structural factors" (Sewell 1992) — as key for understanding how history ultimately unfolds. In contrast, others like Dilthey and Simmel, and continuing through Geertz and other late-20th-century anthropologists, have argued that understanding human societies requires particular attention to individual agency, which, it is claimed, cannot be reduced to impersonal processes (although enacted within the "piled-up structures of inference and implication" provided by culture Geertz 1973: 7). (We are unaware of any proponents for a third view – that history is mostly contingencies, with randomness paramount rather than agency or rules – though Ingold's description of wayfaring as opposed to point-to-point travel seems germane Ingold 2007.) More recently, environmental factors have gained prominence in the study of human history, encompassing both gradual, predictable pressures such as climate change (e.g., Weiss *et al.* 1993, Strawhacker *et al.* 2020) and unpredictable, exogenous events, like volcanic eruptions (Peregrine 2020). In a similar vein, diseases exert influence through continuous pressure (Durham 1991) and as well as episodic shocks in the form of epidemics (Harper 2021).

This brief overview does not exhaust the perspectives of social scientists and historians on what drives history, but it captures many prominent and disparate tendencies. Can we unify these perspectives into an integrative, formal framework for describing and investigating history? Better still, can we construct such a framework that is also well-suited to analyzing the new historical datasets that are rapidly being constructed and made publicly available?

Of course, many social scientists attempt to mediate some of these different perspectives on history by emphasizing interactions among actions by specific individuals and structural contexts – linguistic, material, economic, etc. – within detailed narratives. They often invite us to informally imagine webs or meshes or other entangling relationships that may constrain agency (e.g., Hodder 2012). These approaches, however, are semi-formal at best, and are not readily extended to include fundamental roles for randomness, exogenous perturbations, gradual environmental processes, etc. Moreover, they cannot take advantage of the great analytic power provided by modern methods in machine learning and statistics.

In this paper we consider a different approach, based *on n-order Markov stochastic processes*. As described in more detail below, such a process maps any sequence of n states of a dynamical system occurring at n different times all $\leq t$, to a Gaussian distribution giving the state of the system at any future time $t' = t + \delta$. An important feature of such processes is that the smaller δ is — the less far into the future we are looking — the tighter that Gaussian distribution.

We can view deterministic ("rule-based") dynamics as the limiting case of a Markov process where no matter how far into the future we predict (i.e., no matter how large δ is), the Gaussian distribution has zero width. In contrast, purely random ("contingent") dynamics is the limiting case of a Markov process where the Gaussian distribution governing possible future states of the system is extremely wide, even for times very near in the future (i.e., δ very small). These features suggest that we can gainfully cast the semi-random dynamics of human groups through spaces of social, political, economic, environmental, and other variables — a dynamics that is midway between fully deterministic and fully contingent dynamics — as a (Markov) stochastic process.

Going further, we can readily accommodate discontinuous perturbations to such an underlying Markov stochastic process, simply by using appropriately modified, more complicated stochastic process models like Levy distributions (Lawler 2018). Alternatively, we can *detect* such perturbations in a time-series dataset, as events in that time-series that have low likelihood under some particular Markov process. In both of these ways we can readily incorporate perturbations in our analysis of historical time-series data sets based on stochastic process models, whether those perturbations are environmental, or due to the actions of particularly important individuals.

There are other substantial advantages to using (Markov) stochastic processes to model historical dynamics, in addition to unifying some of the most popular perspectives on what drives the dynamics of history. First, like any other formal framework, adopting a stochastic process perspective forces us to be precise and explicit in our thinking about the dynamics underlying any historical dataset. Second, stochastic process models in particular provide a way to quantitatively investigate the randomness in the dynamics of human groups, how that randomness varies depending on the characteristics of the group, and how that randomness differs from the observational noise inherent in the construction of all historical datasets. Third, stochastic process models allow us to use the data to distinguish when external perturbations to the dynamics of a particular human group actually occurred, and when instead an apparent perturbation was actually the result of purely human processes. More broadly, analyzing historical datasets as stochastic process models allows the data to identify how human social groups evolve, possibly revealing emergent phenomena like unanticipated relations, rather than rely on preconceived notions.[1]

Fourth, as discussed in more detail below, there have recently been some very powerful algorithms developed in the machine learning community that allow us to fit stochastic processes to the time-series datasets commonly encountered in history, archaeology, and related fields. Many such datasets contain large amounts of missing data. Indeed, one of the major challenges with analyzing archaeological time series data using conventional techniques is the need to either "imputate" missing data in the time-series (as in (Turchin *et al.* 2018a)), or simply remove from the analysis any time-series that has too much missing data (as in (Whitehouse *et al.* 2019); see also (Beheim *et al.* 2021, Whitehouse *et al.* 2021)). As we describe in Section IIC below, recently there have been stochastic process estimators constructed that allow us to both avoid the entire issue of imputations, and to use all of our data, without throwing any out.

Furthermore, many datasets encountered in history and archaeology have a large number of variables but relatively few observations. One's statistical method must either be appropriate for such a dataset or utilize a dimensionality reduction, such as principle component analysis (PCA), as a first step (more on this below). There have been some recent advances in machine learning on latent space representations of observations that could prove useful in addressing this problem.

In addition, having an underlying *first-order Markov* stochastic process model of the dynamics of a social group has particular advantages for the social scientist. Such a model tells us how the future dynamics of a human group directly depends on the current characteristics of that human group, i.e., on the current position in a vector space of such characteristics. In contrast, the insights from a high-order time-series analysis of the same dataset are often more opaque, being of the form, "if the social group has had a particular *sequence* of characteristics in a succession of time periods, then its future dynamics is likely to be ...". Similarly, a first-order process tells us how the randomness in future dynamics varies directly as one changes current characteristics.

Furthermore, many historical and archaeological datasets contain very few data in a very large space. This means that a first order Markov model with (far) fewer parameters than higher order models is a sensible baseline model. Indeed, if the data is particularly sparse it may not even be possible to use a higher order model without over-fitting the data[2] Clearly relatively high-order, possibly expert-defined, models are suitable for certain problems (discussed further below). Even in such situations though, first-order models can provide a starting point, and model selection techniques can be employed to determine suitable assumptions.

Summarizing, recent advances in machine learning make using stochastic process models to analyze historical datasets both possible and desirable, and we have outlined some practical advantages for doing so. Current machine learning learning algorithms are improving rapidly. Adopting and modifying state-of-the-art research in machine learning, together with reaching some consensus on best practices for handling missing data and performing dimensionality reduction, will greatly benefit historical and archaeological research.

Roadmap: In the next section we present a high-level summary of stochastic process models, beginning with a discussion of first-order Markov processes, and then discussing higher-order processes.

To ground these abstract considerations, we then present several examples of how to analyze historical datasets with stochastic processes, chosen to illustrate the strengths of stochastic process modeling described above. The first set of examples presents recent research that explicitly involves stochastic process models. Following that, we present several examples of recent research that do not explicitly invoke stochastic process models, but develop analytical approaches or datasets that (we argue) can contribute to the development of such models. We end with a section suggesting possible near-future research projects fully grounded in a stochastic process perspective.

Importantly, the more powerful the tools for analyzing a time series dataset, the greater the opportunity to misuse them. This issue is especially challenging when (as is often the case in the social sciences) the datasets are relatively small, and / or have many missing entries, or entries with some missing values. Therefore one must be quite careful when estimating an underlying stochastic process from a historical dataset. Accordingly, we also use the examples to illustrate some potential pitfalls and how to avoid them. In the same pragmatic spirit, we present lessons learned in many of the examples.

Through these examples, we hope to convincingly demonstrate that the framework of stochastic processes

TIME SERIES TYPE	ALGORITHM
Homogeneous and/or very little data	(Pavliotis 2016)
Inhomogeneous, missing fields in data	(Yildiz et al. 2018)
Inhomogeneous, no missing fields in data	(Kidger et al. 2020, Li et al. 2020)

Table 1 Three type of time-series data over vector spaces, and papers showing how to fit those types of time series with a Markov process. (Current as of 2023.)

is particularly well-suited to uncovering deep regularities in the unfolding of history. More broadly, we argue that a stochastic process perspective provides a way to (at least start to) unify the many perspectives on the unfolding of history promoted by previous researchers.

One of the most exciting aspects of fitting low-order Markov models to time series data is how quickly the field of machine learning is generating advances in techniques for doing this. As an aid to the reader, in Table 1 we highlight some of the more prominent such techniques that exist at present. In that table, "homogeneous" means that the parameters of the Markov process do not vary across the (Euclidean) vector space. As an illustration, suppose that a human society is characterized by a set of n real numbers, which collectively specify a vector space. We would say the evolution of that society is homogeneous if it does not depend on where it is in that vector space. For example, the dynamics of a uniform drift up the diagonal of that vector space is homogeneous. "Fields" refers to the components of such a vector space. Missing fields, ubiquitous in social science data, are data points where one more of the components of each vector data point is missing some of its entries.

The primary focus of this paper is Markov models over Euclidean vector spaces, i.e., vector spaces where every component of the vector is a real number. Accordingly, in the interests of space, we do not present here any of the techniques for fitting Markov models over discrete, finite spaces. This important issue is discussed below.

1 STOCHASTIC PROCESSES

1.1 FIRST-ORDER MARKOV PROCESSES

To make the ideas above more precise, we now introduce the most basic type of stochastic process, which we shall refer to repeatedly in the sections below. This process is called a "(first-order, time-homogeneous) Markov process", with the associated equation for the dynamics sometimes referred to as a "master equation."

Let $\mathbf{x}$ be the location of a society in some appropriate socio-political-environmental state space, and let t be time. Consider the evolution of the society through that space starting at the value $\mathbf{x}_0$ at time t_0. (For the moment we take $\mathbf{x}$ to be a real-valued vector, for expository purposes.) A fixed (time-invariant) stochastic process assigns a probability to each trajectory of $\mathbf{x}$ emanating from this starting point. In a Markov stochastic process, there is a parameter $\boldsymbol{\theta}$ specifying a "transition matrix," or kernel, $\mathbf{W}_\theta(\mathbf{x}'/\mathbf{x})$, and the probability distribution of trajectories through $\mathbf{x}$ is governed by the linear differential equation,

$$\frac{d}{dt}p_t(\mathrm{x}) = \int d\mathrm{x}'[\mathrm{W}_\theta(\mathrm{x}|\mathrm{x}') - \mathrm{W}_\theta(\mathrm{x}'|\mathrm{x})]\, p_t(\mathrm{x}') \tag{1}$$

Intuitively, the term $\mathbf{W}_\theta(\mathbf{x}/\mathbf{x}')$, captures the flow from all states $\mathbf{x}'$ into $\mathbf{x}$, while the term $\mathbf{W}_\theta(\mathbf{x}'/\mathbf{x})$ captures the flow from x into all other states.

As (1) illustrates, in a stochastic process it is not the state of variable $\mathbf{x}$ that evolves deterministically and continuously. Instead, it is its *probabilities* that do. These probabilities tell us how the trajectory of values $\mathbf{x}(t)$ can fluctuate, in general allowing both smooth, drift-like behavior, as well as discontinuous jumps. Importantly, there are two types of such jumps. If the system is currently at $\mathbf{x}'$, and the kernel $\mathbf{W}_\theta(\mathbf{x}/\mathbf{x}')$ places non-infinitesimal weight on an $\mathbf{x}$ that lies a non-infinitesimal distance from $\mathbf{x}'$, then there is nonzero probability of a discontinuous jump across the space. On the other hand, it could be that a small change in the current position, $\mathbf{x}'$, results in a large change in the kernel $\mathbf{W}_\theta(\mathbf{x}/\mathbf{x}')$, and therefore a large change in where the system is likely to move from $\mathbf{x}'$. This can be viewed as a "jump" in the *derivative* of $\mathbf{x}$, arising from a small change in $\mathbf{x}'$ i.e., a jump the "direction of flow" of the system, rather than in the position of the system itself. The parameter $\boldsymbol{\theta}$ parameterizes the stochastic rule, and so directly specifies where in state-space either type of jump occurs.

No matter what the precise stochastic process model being used to explain historical data, we will refer to the average-case, smooth patterns in the trajectories through the space of random variables as the **trends** emphasized by structural approaches. These include patterns in the trajectories through the space of environmental variables as well as through more strictly socio-political ones. Similarly, discontinuous **jumps** in the trajectories of the random variables often but not always involve human agency. The distinction between trends and jumps is often invoked in the historiography of science. For example, the invention of new ideas or technologies can occur through a gradual process, such as with Moore's law. New ideas can also occur through a jump, however, such as Einstein's invention of general relativity.

In addition to such changes in the value of $\mathbf{x}$ at some time t that are governed by the stochastic process, it is possible that the *stochastic process itself* changes at some time t. Such a change in the process would not affect the immediate value $\mathbf{x}(t)$, but rather it would affect how $\mathbf{x}$ is likely to evolve after time t.[3] We refer to such a change in the stochastic process itself as an **exogenous perturbation** of the stochastic process.[4] Formally, an exogenous perturbation is a change at some time t in the parameter $\boldsymbol{\theta}$ in (1), due to an event not captured in the stochastic process variables, $\mathbf{x}$.

To illustrate these points, suppose $\mathbf{x}$ is a vector of variables related to socio-economic characteristics of a particular society. For that space of variables $\mathbf{x}$, slow changes in climate due to reasons independent of human activity constitute gradual exogenous perturbations. For the same space of variables, events like volcanic eruptions would also be exogenous perturbations, but sudden ones. Similarly, the birth of a great leader who is in fact historically consequential (not just idolized that way by future generations) would be a sudden exogenous perturbation. It can be quite challenging to ascertain from a given historical dataset when (if ever) such exogenous perturbations happened, or if instead the dynamics of the data through the space of values $\mathbf{x}$ just reflects evolution under (1) for an unvarying $\boldsymbol{\theta}$. Yet making this determination is crucial if we wish to infer any general principles of historical dynamics. We touch on this issue several times in the examples below, and also discuss it with some actual historical dynamics from the US Southwest in the supplement, Section III A.

The foregoing summary of Markov processes elides many potentially important details. To present some of those details, in the SI Section 1 we work through perhaps the simplest Markov process, a 1-dimensional random walk with step size 1. We also discuss some important variants to (1) there.

1.2 HIGHER-ORDER STOCHASTIC PROCESSES

It is important to emphasize that the first-order Markovian dynamics given in (1) is only one type of stochastic process. If a system is first-order Markovian then the values of our variables in the past provide no useful information beyond what's in their current value for predicting their future values. However, in many common circumstances this property will be violated, at least to a degree, and so the system will not evolve precisely according to (1).

As an example, suppose we coarse-grain the values of $\mathbf{x}$ into large bins, e.g., to help in statistical estimation. Then in general, even if the dynamics over $\mathbf{x}$ is first-order Markovian, the dynamics over those coarse-grained bins will not be. (We illustrate this with the random walk example in the supplement, Section 1) Similarly, if we leave some important components of $\mathbf{x}$ out of our stochastic process model, they become what are called "hidden variables". In such circumstances, again, the dynamics over the visible components of $\mathbf{x}$ will not be first-order Markovian (although it might be n-th order Markovian for $n > 1$).

In all of these scenarios, alternative techniques like Hidden Markov models (see Section II D)), or high-order Markov models, may be more appropriate than first-order Markov models obeying (1). Another example of a powerful alternative technique is delay-embedding time-series analysis (Takens 1981, Sauer *et al.* 1991) versions. In some cases, such use of delay coordinates to predict future trajectories can be formulated as a technique for estimating the average future trajectory under a high-order Markov model.[5] In particular, vector autoregressive moving-average (VARMA) models (Lütkepohl 2005) can be formulated as such a stochastic process.

While such higher-order models have been applied before in historical analyses, they have several non-trivial shortcomings, as discussed above. Nonetheless, there are situations where it is *a priori* plausible that using a higher-order process would provide substantially higher predictive accuracy of the fit to the time series data. Fortunately, recent research in machine learning has developed techniques to to fit a continuous-time Markov process over a "hidden space", with the visible time series being a projection of that Markov process onto a "visible space" (Kidger *et al.* 2020, Li *et al.* 2020). (These techniques are based on cutting-edge deep-learning algorithms like generative adversarial networks, and variational auto-encoders.) Intuitively, these techniques fit the time-series data with a continuous-time hidden Markov model (HMM). In general, such a continuous-time HMM can capture behavior in the dynamics over the visible space that is more than first order. Crucially, these techniques for fitting continuous-time HMMs avoid the curse of dimensionality that plagues conventional approaches to time-series analysis involving delay coordinates. There are also situations when using a continuous-time HMM is inappropriate, and a conventional discrete-time HMM makes more sense. Section IID below discusses an example of this situation in detail.

To keep this perspective article focused however, aside from Section IID, in the parts of the discussion where we explicitly consider Markov processes, we will concentrate on first-order Markov processes, without hidden variables. In other parts of our discussion we won't even explicitly specify the precise type of stochastic process we are considering. In these parts of the discussion we analyze the stochastic process at an even more abstracted level (e.g., in Section IIC).

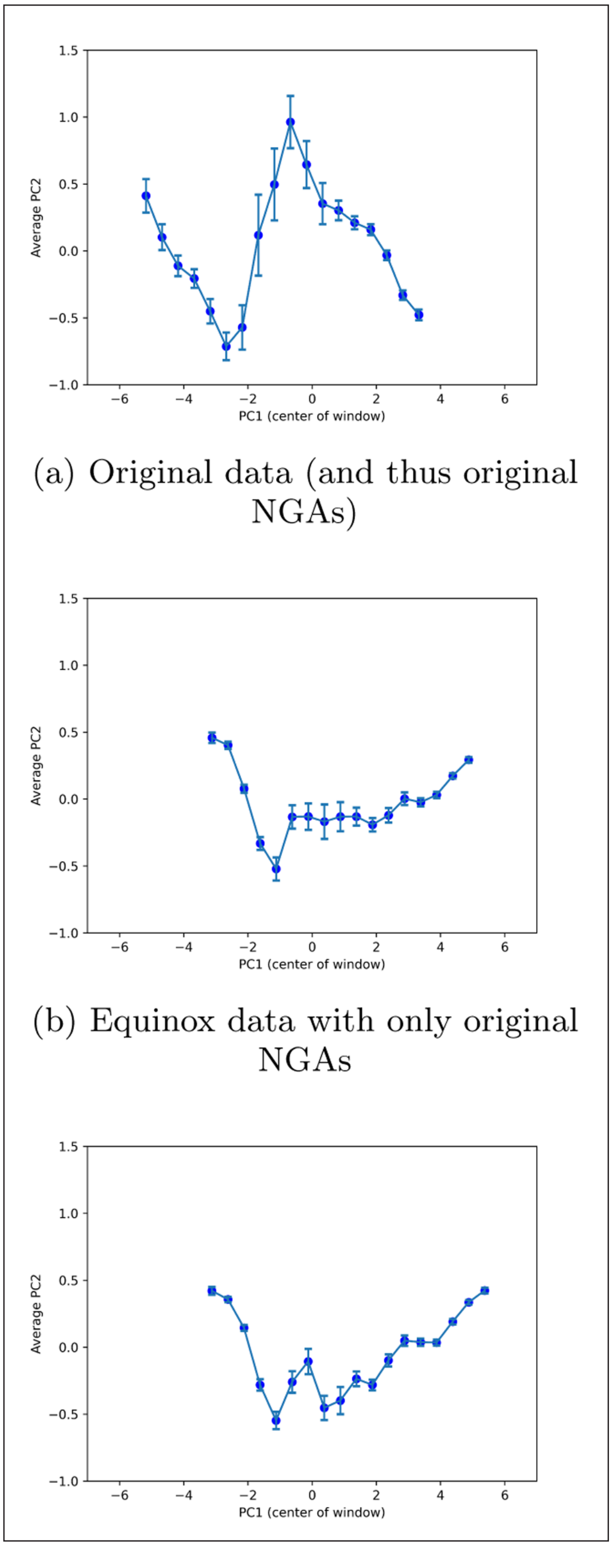
2 RECENT RESEARCH IN HISTORY BASED ON STOCHASTIC PROCESS MODELING

2.1 JUMPS IN SOCIOPOLITICAL COMPLEXITY OF POLITIES IN THE *LONGUE DURÉE*

It is widely agreed that over long-enough periods of time there is a strong trend for polities to increase in size, or to disappear (e.g., Tainter 1988). Less clear is whether such increases are relatively smooth trends in which various measures of scale and/or capability, such as information processing or capability at warfare, increase in lock step; or whether there are clear discontinuities or disjunctions in the growth processes.

The Seshat dataset (Francois *et al.* 2016) allows analyses that can address such questions. This dataset contains values for a large and diverse set of variables concerning societies over the past several thousand years, in both Afro-Eurasia and the Americas. (We provide more background in the supplement, Section 3.) This dataset has been condensed for some recent analyses into a set of nine complexity characteristics (CCs) defined by the researchers for multiple past societies. These are reported at (usually) one-hundred-year intervals for the dominant polity controlling each of thirty world regions called Natural Geographic Areas (NGAs; the controlling polity may have its capital outside the NGA). For some NGAs, these time-stamped data stretch back millennia. PCA on the CCs found that the primary component, PC1, explained roughly three-quarters of the variability of the 9 CCs (Turchin *et al.* 2018a). The Seshat team suggests interpreting PC1 as an overall, scalar measure of social complexity (Turchin *et al.* 2018a, Whitehouse *et al.* 2019) This is because, by construction, the individual CCs represent separate complexity measures, and PC1 "captures three quarters of the variability". Note as well that PC1 has a roughly equal loading on all 9 CCs – that is, it is a weighted version of the individual complexity measures. Note that PCA is independent of the time-stamps of the individual data points. So mathematically, there is no reason to expect that PC1 bears any relation to time; it could just as readily decrease with time as increase. Nonetheless, examination of the data makes clear that there is a strong tendency for polities (and sequences of polities within an NGA) to increase their PC1 value with time, albeit with long periods of stasis (Turchin *et al.* 2018a). Crucially, different NGAs have different starting dates, and there is no guarantee that sociopolitical change will march in lock-step across NGAs. On the other hand, as discussed below, PC2 consists of two sets of CC components: one related to polity scale and another related to information processing capabilities (Shin *et al.* 2020). This suggests fitting a very basic "stochastic process model" to the Seshat data, by graphing PC2 against PC1, with larger PC1 values generally corresponding to later times in an NGA's sequence.

A recent paper (Shin *et al.* 2020) explored this possibility by analyzing how PC2 and PC1 co-vary. Whereas PC1 has positive loadings on all 9 CCs, PC2 has negative loadings on the scale variables (such as polity size) and positive loading on information variables (such as writing, texts, and money). Interestingly, as illustrated in Figure 1, PC2 is a "saw-tooth" function of PC1, first decreasing until PC1 is about –2.5, then increasing until



(a) Original data (and thus original NGAs)

(b) Equinox data with only original NGAs

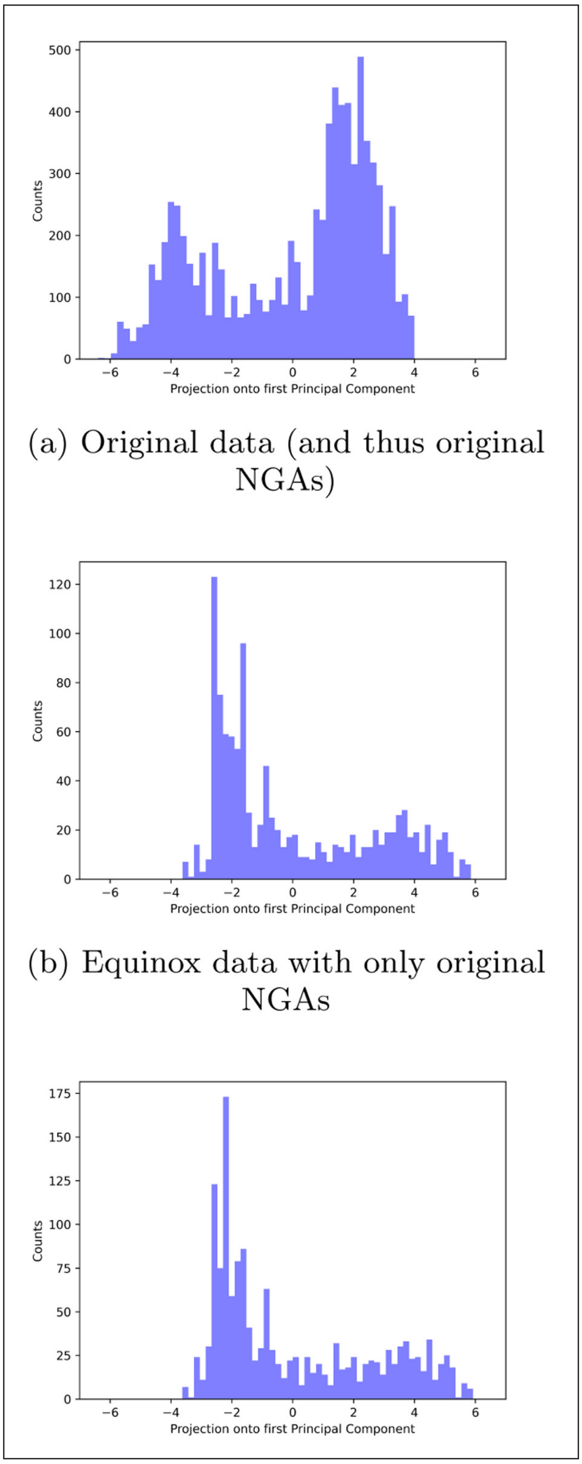
Figure 1 The mean value of PC2 as a function of PC1, where the mean and error bars are calculated using a sliding window with a width of 1.0 in PC1-space. The top curve is identical to that in (Shin et al. 2020). The middle plot uses the new Seshat Equinox dataset, but the new NGAs added to Equinox have been removed. The bottom plot uses the new dataset and utilizes all Equinox NGAs. The key difference between the original datasest and the Equinox dataset is that the latter dataset uses one fewer CCs, collapsing the two original information processes CCs into a single CC. These figures can be reproduced using the code in the publication github repository (see supplement).

PC1 is about –0.5, then decreasing again. This reveals an unexpected sociopolitical phenomenon: there is a strong average tendency among the polities in this sample to first increase in size until a certain threshold size is reached (PC1 = –2.5), at which point the dynamics is taken over by increases in the information-processing ability of the polity (broadly defined) with relatively little increase in size. Eventually this process reaches a new threshold (PC1 = –0.5), after which a dynamic process increasing the size of the polity again takes over.

The preceding remarks apply to the original dataset analyzed by (Shin *et al.* 2020) and (Turchin *et al.* 2018a). An expanded version of the Seshat dataset called Equinox (Turchin *et al.* 2020) has recently been released that displays some different patterns. There are four primary differences between the original dataset and the Equinox dataset. First, some of the entries have been modified, for example, to fix errors. Second, four new NGAs have been added. Third, the Equinox version contains only one imputation for every observation, as opposed to twenty for the original dataset. Fourth, the Seshat team has collapsed the original 9 CCs into 8 CCs. This choice and the associated methodology does not appear to be documented anywhere, but it seems that the two original information processing CCs (Writing and Texts) have been collapsed into a single CC (Info). This collapse yields a qualitatively different relationship between PC1 and PC2, which is perhaps not surprising since information processing is a crucial consideration in (Shin *et al.* 2020). To determine the primary driver of the qualitative differences, we reproduced the sliding window plot for the new Equinox dataset using only the original NGAs and with all NGAs (the middle and bottom plots of Figure 1). Because the two plots appear very similar, it does seem that the primary driver is the CC collapse.

We have also plotted histograms (Figure 2) and movement plots (supplement Figure 3) for each of the three cases. While the updated sliding window plots pick up the same two hinge points as in the original Seshat data, they differ markedly for larger values of PC1. Similarly, the means and variances of the two components of the PC1 histogram may be similar, but the component with the larger mean has lower density (i.e., the weighting of the second mixture is lower). Pending a more detailed analysis of the differences, we believe the main message from this comparison to be cautionary: variable choice and methodology used for dimensionality reduction (or to otherwise simplify the dataset) matter. This is a topic that is largely outside the purview of this perspective piece, but should, we think, be a major topic for the discipline as a whole.

A key point shared by these two analyses (portrayed admittedly more clearly in the first) is that values along PC1 at approximately –2.5 and –0.5 appear to constitute *thresholds in complexity*, with different sociopolitical processes dominating the dynamics depending on whether the PC1 value is less than –2.5, between –2.5 and –0.5, or greater than –0.5. This tendency is exhibited by polities that reached these thresholds at vastly different times, in far-separate locations and differing



(a) Original data (and thus original NGAs)

(b) Equinox data with only original NGAs

Figure 2 Histogram of PC1 values for all polity-time pairs. The top curve is identical to that in (Shin et al. 2020). The middle plot uses the new Seshat Equinox dataset, but the new NGAs added to Equinox have been removed. The bottom plot uses the new dataset and utilizes all Equinox NGAs. The key difference between the original datasest and the Equinox dataset is that the latter dataset uses one fewer CCs, collapsing the two original information processes CCs into a single CC. These figures can be reproduced using the code in the publication github repository (see supplement).

local environments, with extremely different neighbors. Therefore these two thresholds cannot be due to exogenous perturbations, in which $\boldsymbol{\theta}$ would change at a specific time. Likely they are instead examples of jumps of the second type discussed above, where the transition matrix $\mathbf{W}_{\theta}(\mathbf{x}'|x)$ in (1) changes discontinuously once $\mathbf{x}$ (in this case, PC1-PC2) reaches a particular threshold.

One benefit of stochastic process modeling is that it fosters cumulative cascades of analyses, in which an insight from analyzing one dataset allows us to analyze a second set in a way that would not have been possible otherwise. We can illustrate this with the current controversy about the relationship between when a polity undergoes a jump in "social complexity", and when its members widely adopt worship of "moralizing gods." One side of the debate are researchers who say that the emergence of such gods is a precondition for a jump in social complexity. (Loosely speaking, the reasoning underlying this hypothesis is that worship of such deities fosters wide-spread sacrifice in pursuit of the common good, which in turn results in the jump in social complexity.) Others have argued just as vigorously that the causal influence is the other way around, that such jumps are actually a necessary condition for the emergence of moralizing gods in a society.

We can exploit our analysis described above concerning hinge points to address this issue. To begin, we suppose that the first hinge point is a marker for a jump in social complexity. As described in App. A(2), we can then analyze the relationship between the time that moralizing gods appear in a polity and the time when that polity undergoes a jump in social complexity. The results contrast with the somewhat polarized views in the literature (Whitehouse *et al.* 2019, Beheim *et al.* 2021, Purzycki *et al.* 2016). Given a null hypothesis that polities with moralizing gods arise later than polities without them, there appears to be little evidence of correlation between the time that moralizing gods appear in a polity and the time that that polity experiences a "jump of social complexity".[6] Such gods do not seem to be a precondition for jumps in complexity by a polity, nor does it seem that a polity's having gone through such a jump is an absolutely necessary precondition for their appearance (as claimed in (Whitehouse *et al.* 2019), now retracted by many of its authors).

This is only a suggestive, preliminary analysis, not meant to be definitive. Historically, its only use was to cast doubt on the claim in (Whitehouse *et al.* 2019), before the controversy about that paper arose which resulted in its retraction. A natural way to extend this preliminary analysis would be to fit a full stochastic process model to historical data in a state-space $\mathbf{x}$ that consists of both the CCs of a society (or one or more of the corresponding PC-values) and a variable representing the presence/absence of moralizing gods.

2.2 ARE THERE STABLE SOCIAL EVOLUTIONARY TYPES OF SOCIETIES?

Social researchers sometimes argue that some set of multiple time-series across a space seems to move quite slowly across one or more regions of that space. This has sometimes been interpreted as meaning that those regions are "basins of attraction" or "attractors" or "equilibria" of an underlying stochastic process (Peregrine 2018). A long-running interest in archaeology (stretching back at least to the mid-late nineteenth century (Tylor 1871)) is whether there really are such stable social evolutionary types of societies, or whether any apparent instances of such societies in the data are largely illusory. Tendencies to see such modes seem to be strongly influenced by fashions in research; the fashion for several decades has been to minimize their existence.

In this subsection we illustrate how careful analysis of stochastic processes can help address this issue empirically, again using the version of Seshat dataset analyzed in (Turchin *et al.* 2018a). If one plots a histogram of the PC1 values of all the time-stamped observations across NGAs, one finds that, even though every polity is usually sampled once per century (though this is less true early in time for polities with long time sequences), the PC1 values cluster into two widely separated regions (see Figure 2). This *could* imply that those regions of PC1 values are "equilibria" of the underlying dynamics, i.e., stable social evolutionary types—but is such an inference warranted?

While some have thought so (Miranda and Freeman 2020, Turchin *et al.* 2018b), an analysis using the sorts of tools we advocate in this paper suggests that the answer is no. Specifically, those clusters can be explained as arising from a null model that results from the interplay between the starting values of polities in the dataset (i.e., the point at which they are first coded), the underlying stochastic process, and the sampling of the stochastic process. To give a simple, counter-intuitive example of that interplay, even if a stochastic process has the same average speed across a space, if the *variance* of its speed varies across that space, then we will see clusters; regions with higher variance will have clusters of more data points than regions of lower variance (Shin *et al.* 2020). Complicating the picture still further is the fact that the Seshat dataset comprises multiple, different time-series, reflecting a combination of three factors:

- **a)** Random variation across the time series of the polities in the PC1 value of the earliest data point;
- **b)** Random variation across the polities in the chronological time of that first PC1 value;
- **c)** A general drift of polities from low to high PC1 values that is well-modeled by a time-homogeneous Markov chain.

Perhaps the most conservative way to test whether such factors lead to features like clusters in the dataset is to use a null hypothesis that the data were generated by a first-order, time-*homogeneous* discrete-time Markov chain, corrected so that the variance of the chain as one moves across the space matches the variance in the dataset as one moves across the space.

Indeed, as described in the supplementary materials of (Shin *et al.* 2020), one can accurately fit a first-order, time-*homogeneous* discrete-time Markov chain to the Seshat dataset PC1 values, mapping one PC1 value to the next with a conditional distribution that does not change in time (as it ought, if polities were "dwelling" in certain portions of its space). If one forms 30 sample time-series by running that Markov chain 30 times, each for a different number of iterations, and superimposes the 30 resultant time-series, one finds that they will typically form two widely-separated clusters, just like the clusters in the original Seshat data along the PC1 dimension (Figure 2). Yet because the underlying process is a time-homogeneous Markov chain, there are *no* attractors in the underlying dynamics.[7]

Intuitively, these clusters reflect several factors. First, the 30 time-series are all transients (none long enough to be samples of the stationary distribution of the Markov chain). Second, they are all of different lengths (due to (a, b) above). Combined with the fact that the generating conditional distribution is non-uniform, the result is that two clusters in the data appear, with no direct significance for interpreting the underlying dynamics. We discuss this example in some detail to highlight the formal and theoretical subtleties in using stochastic processes to model archaeological datasets. But, as this analysis shows, these challenges can be met—supporting our claim that such a program can now be developed successfully.

2.3 COMPARING DIFFERENT HISTORICAL TRAJECTORIES: FLOW MAPS

Another big question in history is to what degree different trajectories are comparable. Formal analysis of the sort presented here allows us to move beyond metaphorical uses of developmental stages or historical cycles. In Section IIB we presented a cautionary tale concerning computational history, arguing semi-formally that a simple null-hypothesis test shows that the Seshat dataset is consistent with the hypothesis that it was generated by a time-homogeneous Markov chain. While such null-hypothesis tests are a useful starting point, they are blunt instruments that sidestep the critical question of what is the best statistical inference from the data.

Fortunately, as mentioned in the introduction, recently developed analytical tools provide very powerful ways to estimate first order Markov processes from time series datasets (see (Yildiz *et al.* 2018, Kidger *et al.* 2020, Frishman and Ronceray 2020, Friedrich *et al.* 2011, Hoffmann *et al.* 2021, Davis and Buffett 2022, Craigmile *et al.* 2022) for a representative sample of recent papers). These promise to be extremely useful for investigating historical datasets. In this section we illustrate how one such recently developed tool (Yildiz *et al.* 2018) can produce a far more sophisticated estimate of the stochastic process that generated the Seshat data than a simple fit with a time-homogeneous Markov chain.

Figure 3 presents a plot of trajectories of the polities in the Seshat dataset in the joint PC1-PC2 space. Each of those 30 time-series is quite short, which makes it difficult to extend any single one of them, e.g., by using vector autoregressive-moving-average (VARMA) models or delay coordinate techniques (Lütkepohl 2005). A different approach is to model the dynamics across PC1-PC2 by fitting the data to a Markov process as in (1), with its transition matrix restricted to the set of stochastic processes called "Langevin equations" (a.k.a. "stochastic differential equations" [SDEs]) where the drift and diffusion terms vary across the space. As an illustration, Figure 3 shows the drift term of the Langevin equation given by applying the Bayesian estimation technique recently introduced in (Yildiz *et al.* 2018) to the time-series across PC1-PC2 in the Seshat data.

There are several advantages to this kind of fit of an SDE for sets of short time series. Most directly, as discussed above, the fact that PC1 explains most of the variability of the Seshat dataset by itself provides no basis for identifying the PC1 position of a polity as a measure of its social complexity, despite the fact that it has often been explicitly used that way (Whitehouse *et al.* 2019). One could indeed argue that any dataset,

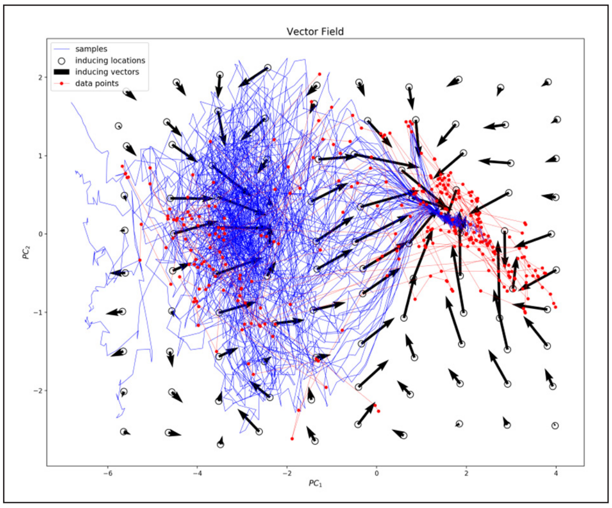


Figure 3 The two axes are PC1 and PC2 of the (original) Seshat dataset, respectively. The red dots are the elements of that dataset. Since those elements of the dataset are time-stamped, one can find the Langevin SDE that has highest posterior probability conditioned on that dataset. The black arrows are the mean velocity vectors of that Langevin equation at the associated PC1-PC2 positions (i.e., they are values of the drift vector field of that SDE evaluated on a grid). Finally, the blue lines are counterfactual, sample trajectories of that SDE.

by itself, cannot provide guidance as to how to quantify "social complexity" given the vague and intuitive nature of the concept, which tends to mean different things to different people.

In contrast, a flow field like that in Figure 3 focuses attention on the more concrete issue of how real-world societies actually evolve. For example, it has long been noted that some sufficiently coarse-grained datasets seem to suggest that there is a rough tendency for polities to go through historical cycles (Khaldun 2015). As future work, one can imagine estimating the SDE that generated those datasets, then employing Monte Carlo sampling of that SDE to quantify this tendency of polities to go through such cycles. For example, this could be done by Monte Carlo estimating the probability that a polity currently at a given $\mathbf{x}$ will follow a trajectory that goes at least ϵ_1 away from $\mathbf{x}$ but then returns to within $\epsilon_2 < \epsilon_1$ of $\mathbf{x}$, as a function of $\mathbf{x}$, ϵ_1 and ϵ_2. (In this regard, note the suggestion in Figure 3 of a circulation pattern around the point (1.5, 0)).

In addition, it may prove fruitful to re-express the stochastic flow field (or more precisely, the field of drift vectors defining the SDE). In particular, robust algorithms have been developed in computer graphics for inferring a Helmholtz decomposition (Arfken and Weber 1999) of such flow fields. Such a decomposition expresses the flow across PC1-PC2 in terms of (the gradient of) a scalar potential field and (the curl of) a vector potential field. The scalar potential field might prove particularly illuminating; level curves of that potential field across PC1-PC2 would be an extremely flexible way of identifying polities all of whom are at the same "stage of development". In the absence of a vector potential, the dynamics of polities would simply be gradient descent, descending those level curves.

At a high level, this joint use of tools for inferring SDEs and the Helmholtz decomposition is a way to infer directly from the data novel *laws* governing social systems (or at least strong empirical regularities emerging from them), formulated in terms of the Helmholtz decomposition. Note though that both the SDEs and the resulting potentials might not have a simple, explicit form. That could make it difficult to give them a simple social science interpretation. This difficulty is actually fundamental in all analysis of longitudinal datasets; in particular a similar interpretational challenge holds even for fitting one-dimensional time series datasets using delay coordinate techniques, even linear ones like ARMA models (Takens 1981, Sauer *et al.* 1991). Addressing this challenge is a topic for future research.

2.4 HIDDEN MARKOV MODEL STOCHASTIC PROCESSES

While archaeologists often debate which types of variables are most important for understanding a given region's past, a few of these are often considered more crucial than others, including climate (principally temperature and rainfall), demography, political organization, technology, economy, ideology, and local ecology. (Clearly all of these could be further deconstructed in various ways.) In this section, we describe in some detail a stochastic process model that can be used for demographic modeling and which can be used to fuse multiple types of data into a single inferential framework. Aside from being potentially useful in its own right, we offer the model as a template for using stochastic process models in archaeological or historical work.

Of the variables related to demography, perhaps the most crucial is a region's total population size, but it is often necessary to consider in addition how that population is structured by age, sex, space, or socioeconomic status. High-quality demographic models exist for describing structured populations (Rogers 1966, Caswell 2001), so the main challenge is how to use demographic models to infer past demography from the available archaeological data.

One complication is the temporal fidelity at which inference is possible. Seasonal and yearly changes in climate can change demography at yearly and sub-yearly timescales, but it is rarely, if ever, possible to infer both climate variability and demography at these timescales. As described in the introduction, this means that the data are coarse-grained, and so might not be accurately described by a Markov process. Therefore a (discrete-time) HMM could prove more suitable. In App. D we describe in detail an age- and sex-structured population projection model. The fundamental demographic variable is the population vector $\mathbf{z}_t$ of females in the population, the elements of which are the number of females in each distinct age class. The time-dependent population matrix projects the population to the next, distinct time-step per

$$\mathbf{z}_{t+1} = \mathbf{A}_t \, \mathbf{z}_t \tag{2}$$

In addition, a corresponding population vector of males exists, which depends on the sex ratio at birth (SRB) of males to females and male age-specific mortality (details in the supplement), which is rarely equal to female age-specific mortality and may fluctuate more, for reasons like warfare.

Substantial work in historical demography and life history theory places constraints on the plausible values of $\mathbf{A}_t$ (Wood 1998, Jones 2009, Jones and Tuljapurkar 2015), or (if adopting a Bayesian approach) can be used to specify priors on $\mathbf{A}_t$. To allow for this we assume there exist $n = 1, \ldots N$ distinct types of annual, reference population projection matrices. For example, one type could be appropriate during a famine or other stressing event, while another corresponds to a stable population size in which high mortality (especially high juvenile mortality) is balanced by high fertility. Let $\mathbf{p}_t$ be a probability vector of length N that gives the probability of being in a given reference state. Transitions between these hidden states occur per

$$\mathrm{p}_{t+1} = \mathrm{W}_{\boldsymbol{\theta}}\, \mathrm{p}_t, \quad (3)$$

where the transition matrix $\mathbf{W}_{\boldsymbol{\theta}}$ depends on a slowly changing, exogenous climatic variable $\boldsymbol{\theta}$, which we propose modeling as a binary or categorical variable that only occasionally changes given the fidelity at which climatic reconstruction is possible (hence, the Markovian assumption applies only so long as $\boldsymbol{\theta}$ is fixed).

What is appealing about the preceding framework is that it allows a diversity of data to be used to infer $\mathbf{A}_t$ and $\mathbf{z}_t$. This has immense value for two reasons. First, it is usually better to use more data (and never worse, if the statistical model is good). Second, while any given type of data is likely biased in distinct and predictable ways (e.g., skeletal collections usually have too low a proportion of the very young, whereas very old individuals are incorrectly aged), these biases are usually different for each data type and uncorrelated across samples, so inference will be far better when multiple types of data are used and sample sizes increase.

This model could be especially valuable if applied to archaeological case studies for which a diversity of factors and variables have been linked. A pertinent example is the so-called Classic Maya Collapse. Between about AD 750 and 1000, various parts of the Maya cultural region in Mexico and Central America experienced a breakdown of sociopolitical systems. Associated with this were major demographic changes (migrations and, likely, overall population decline), changes in long-distance trade routes, intra-elite competition, warfare, and environmental degradation, all in the context of severe drought (Webster 2002, Aimers 2007, Kennett *et al.* 2012, Turner and Sabloff 2012, Hoggarth *et al.* 2017). Exactly how these factors are linked remains an open question, yet a stochastic process model would likely help make these linkages explicit were it to be correctly fitted to the data. For example, drought has been proposed as the major, causal factor leading to the breakdown of Mayan sociopolitical systems, but others argue that all these factors, and more, played some role; using our proposed modeling framework would help elucidate the causes.

3 PREVIOUS STUDIES RECAST IN TERMS OF STOCHASTIC PROCESS MODELS

3.1 INTEGRATING DIFFERENT DATA TYPES FOR NEW INSIGHTS

Like demographic processes, environmental processes have long been seen as catalysts of past social change. In particular, environmental processes are central to research on the history of human-centered food webs (Crabtree *et al.* 2017a, Dunne *et al.* 2016). The benefits of bringing a stochastic process perspective to this research can be illustrated with the investigation in Crabtree et al. (Crabtree *et al.* 2019). That research focused on the question of why small-bodied mammals went extinct in portions of Australia *after* people were removed from the Western Desert.

In the absence of humans, animal populations are often modeled as evolving via Markovian processes (Meyn and Tweedie 2009). However, (Crabtree *et al.* 2019) demonstrated that the extinctions experienced in the Western Desert would not be predicted from the natural baseline extinction rates of such a process. Adopting the perspective of this paper, this result highlights an important modeling issue: did the extinction of small-bodied mammals reflect a change in a hidden variable of an underlying Markovian process, or was it due to a change of a parameter of such a process, via an exogenous perturbation? This distinction is critical because there are statistical techniques for predicting the evolution of Markovian processes with hidden variables, such as time-series analysis, described above. It might be possible with such a model to predict a change in the dynamics of animal populations by looking at time-series data of their dynamics. However, by definition, if the change in the dynamics was due to an exogenous perturbation, then it cannot be predicted from any time-series data, no matter how exhaustive.

In this particular case, it seems unlikely that the change in the dynamics of the animal population level $\mathbf{x}$ can be accurately modeled as the change in the value of a hidden variable, e.g., by using a time-series model of the dynamics of $\mathbf{x}$. In other words, it seems likely that the change in the dynamics of the animal populations reflects an exogenous perturbation of the parameter $\boldsymbol{\theta}$ governing those dynamics. However, this has yet to be formally established. In addition, a more elaborate stochastic process model might expand $\mathbf{x}$ to include some variables concerning the people who eventually forced aboriginal peoples to leave. In such a model, changes to the dynamics of animal populations might reflect the "second type of jump", described above in the discussion just below (1). If that were determined to be the case, it would mean that future values of $\mathbf{x}$ might be predictable, and therefore particular future values of the animal population level might be predictable.

As another example, Yeakel et al. (Yeakel *et al.* 2014) leveraged Egyptian art by coding the taxa present from the Narmur Palette and other datable art work to model animal species extinctions over time. By combining these data with contemporaneous climate data, they found that aridification pulses in the context of a growing Egyptian population played an important role in destabilizing the ecosystem. In the language of stochastic processes, if we take $\mathbf{x}$ to be the combination of human and animal population levels, we see a trend as growing human populations exerting increasing pressures on the ecosystem. These pressures were exacerbated by exogenous perturbations in the form of a series of aridification pulses and a worsening climate.

This example suggests that it may be possible to look at the dynamics of human and animal populations in other regions and infer when aridification pulses, or indeed other climatic events, likely occurred, by testing for exogenous perturbations to the trends in the dynamics of the joint human/animal population levels. In the original context of ancient Egypt, where we already have identified *some* exogenous perturbations in the form of aridification pulses, we could perhaps test for the presence of *other* exogenous perturbations, e.g., due to invasions, or perhaps even due to causes currently absent from the historical record. More broadly, a stochastic process framework could suggest a new avenue for research for Egyptologists: the impersonal trends of extinction and exogenous perturbations of aridification likely had feedbacks on Egyptian society, from impacts on hunting to impacts on cosmology. Exploring how the societal impacts at a given time varied depending on whether there was only an underlying trend or also an exogenous perturbation at that time could lead to new insights on the dynamics of Egyptian social structures.

3.2 STOCHASTIC PROCESSES OVER THE STRUCTURE OF LANGUAGE AND OVER HOW LANGUAGE IS USED

Stochastic processes that are themselves superpositions of Markovian time evolution and branching processes describing the temporal evolution of features in systems that from time to time break up into parts that evolve (largely) independently thereafter. Such processes underlie the history of biology and human languages and are the base of phylogenetics. Data on cognates across many languages, together with a small number of historical anchor points, allows for the reconstruction of language trees and assign approximate dates of divergence. A careful, quantitative analysis of Polynesian languages, for instance, revealed some periods of "stasis" in the history of subfamilies that correspond to distinct phases in the settlement history of the Pacific Islands (Gray *et al.* 2011). Similar arguments have shed light on the expansion of Indo-European language families. Analogous reconstructions of historical population distributions have been conducted using human genetic markers. The idea to integrate genetic, linguistic, and archaeological data is of course not new (Scheinfeldt *et al.* 2010), although so far this has been done at the level of interpreting the data rather then integrating them into a common process model. The latter would be desirable in particular because linguistic changes are strongly impacted by contact phenomena such as borrowing, and even correlate with extra-linguistic factors such as the prevalence of agriculturally produced food (Blast *et al.* 2019). At present, quantitative studies in linguistics are almost always based on datasets that are the result of extensive manual curation — although modern methods in natural language processing might be suitable or at least adaptable to tap into much larger resources (Bhattacharya *et al.* 2018).

Modern large-scale digitized historical archives often provide high-resolution data on the words that individuals—often, though not always, members of a polity's elite—were writing and sharing with each other. Tools from natural language processing are now both simple enough and robust enough to allow social and political scientists to track the flow of complex patterns of language use that correspond to concepts and habits of thought.

Ref. (Barron *et al.* 2018), for example, used the French Revolution Digital Archive (https://frda.stanford.edu) to show the different roles that members of the French revolutionary parliament played in introducing, sustaining, or rejecting novel ideas in the speeches of that country's constitutional debate. These ideas were discovered in an unsupervised manner; rather than pre-determining a list of important ideas, and words that corresponded to them, topic modeling automatically extracted word patterns that could then be back-validated on the speeches themselves. Discovered topics include, for example, concepts as fine-grained as "the possibility of enemies of the revolution within the military".

3.3 ANALYZING THE GREAT ACCELERATION

We are currently living in the Anthropocene. One of the more striking characteristics of this period has been the exponential growth in a large number of important metrics in earth and social systems since the beginning of the 1950s. These growth dynamics are often collectively referred to as the "Great Acceleration" (Steffen *et al.* 2004). What *caused* these transformations is less clear however.

All the metrics that have contributed to the Great Acceleration are, however, dependent on population size, which itself has been changing during this period, to no small degree as a consequence of growth in these metrics. This suggests we fit our data concerning those factors with a stochastic process model, to gain quantitative insight into their joint evolution (see appendices for examples of such a model). In particular, because of the specific growth patterns of the variables, it is natural to fit the data to a stochastic process over the *logarithms* of the variables. When applied to historical datasets, this variant of stochastic process models is known as **historical** or **temporal scaling analysis**. Not only can it give us insights into the joint dynamics of population size and the metrics considered by researchers of the great acceleration. In addition, when performed for different temporal intervals, it gives us insight into *changes* of those joint dynamics. This then invites us to investigate whether those changes in the dynamics are due to exogenous perturbations and/or due to changes in hidden variables (as in the analysis of HMMs in Section IID and/or due to the system reaching a new point of its state space (as in the case of hinge points, considered above in Section II A).

To make this more concrete, suppose that we find that the scaling coefficient remains stable across time. That would be evidence for a simple trend of the underlying stochastic process. Suppose, as an alternative, that the scaling coefficient changes over time. This is evidence for one of three phenomena: either a concurrent change in some exogenous factor driving the dynamics (like a change in climate, or a major drought, or the Green revolution of agriculture); a concurrent change in a hidden socio-political variable generating the observed historical dynamic (like a major war, a world-wide depression, or changes in financial regulation) or simply the system reaching a critical threshold — a new part of the socio-political space — in which the dynamics are different. This illustrates how the stochastic process perspective could invite a host of new investigations, if it was found that the scaling coefficients changed over time.

A recent analysis of the dynamics of the Great Acceleration (Painter *et al.* 2020) determined that socioeconomic, technological, information, biological and earth systems, when scaled to population size, sort into four distinct growth patterns. These patterns suggest fundamental differences in the underlying network of interactions causing the observed patterns of change. Besides the already known sub-linear, linear and super-linear patterns we detected a novel fourth pattern for the Great Acceleration, with scaling coefficients larger than 3, which is unusually large. This pattern applies to parameters that are no longer limited by person to person interactions such as those related to knowledge, technology and finance. These parameters can be interpreted as the main drivers behind the ever accelerating growth dynamics of the Great Acceleration. Decade by decade comparison of these parameters also revealed patterns of change during this time period that correspond well with the possible sources of change that can influence the dynamics of an underlying stochastic process.

Whether an event or a discovery (such as those contributing to the Great Acceleration) is considered a transformative innovation or just a random occurrence affecting a stochastic process can often only be decided in hindsight. What can be done, however, is to asses to what extent such events were surprising or predictable, given the context of the times–here represented by an underlying stochastic process model. Emerging machine learning methods, especially deep learning neural networks, have succeeded at dramatically improving the prediction of quantities across a wide range of contexts. These models have begun to enter the social and historical sciences as complex event and trend prediction (Bainbridge 1995, Davidson 2017, Carley 1996, Zeng 1999, Maltseva et al. 2016, Zhan et al. 2018, MacLeod et al. 2016, Baćak and Kennedy 2019). We can apply this methodology to a number of different events, such as a regime change, adoption of a technology or the emergence of an institution with a particular function (e.g., urban garbage collection). Prediction targets can also be more complex and specific. For example, they have been applied to predict the location, amount, denomination, and material composition of an archaeological cache of money and precious metal hoards, and the distribution of mints from which the coinage derives. This might only be limited by the area, such as surrounding the Mediterranean, and a period of time to which the hoard date (see appendices for a more in depth discussion of these methods).

4 FUTURE WORK AND ASPIRATIONAL CASE STUDIES

4.1 COMBINING TIME SERIES OF HUNDREDS OF RANDOM VARIABLES

As far more datasets become available we will be confronted with combining multiple time series of systems evolving from up to hundreds of different variables, including not only the sorts of variables recorded in Seshat, but also completely different kinds of variables, like lead levels in glaciers, or tree ring widths as climate proxies, to large-scale characteristics of polities that others have discussed before. We will be combining these with wholly different kind of time series, including things like word usage patterns in historical documents, structures in legal codes, time series of fashions, ecosystem characteristics, etc., all concerning different random variables.

As always, great insights will accrue from fine-grained analysis, in which domain experts deeply scrutinize only a few of these time series at once. However, there is also the great allure of integrating all of these myriad time series in a systematic, statistically principled way, to uncover unanticipated connections and insights. Indeed, ultimately, one would want to be able to integrate all of these time series into a single underlying stochastic process, with associated error bars. However, there are many analytical challenges to doing that, due to the sheer number of these times series, and the sheer breadth of the types of associated random variables.

As an important example of such a challenge, how does one infer causalities in a statistically principled way among hundreds of time series, all involving different random variables? For example, what techniques would allow us to uncover statistically significant causal connections relating time series ranging over hundreds of different spaces? Can we scale up techniques like Granger causality from econometrics (Granger 1969) or transfer entropy and directed information from information theory, to such large numbers of spaces (Schreiber 2000, Amblard and Michel 2013)?

As another example, can we extend breakpoint analysis / change detection to involve multiple time series, in order to find non-stationary breaks in the underlying dynamic process? Such breaks would reflect some exogenous events, perturbing the underlying dynamics of the system. So, for example, such a breakpoint might indicate that a particular leader of a state was indeed consequential, rather than just being a "product of their time", in that they perturbed the underlying dynamics of sociopolitical processes — we might conclude from such a breakpoint that a leader had literally changed the course of history.

4.2 LEVERAGING AGENT-BASED MODELS FOR STOCHASTIC PROCESS MODELING

Agent-based models (ABMs) are becoming increasingly instrumental for investigating processes that leave limited physical traces in historical and archaeological data, such as work examining the dispersal of hominins out of Africa (Romanowska *et al.* 2017). Typically such models involve the dynamics of variables that are hidden, in that they do not directly appear in the historical data. For example, Crabtree and colleagues (Crabtree *et al.* 2017b) and Kohler and colleagues (Kohler *et al.* 2018) built a model for sociopolitical evolution in the North American Southwest in which territorial groups undergoing population growth may succeed in subjugating or merging with other groups via warfare or intimidation. Networks of flows of maize as tribute are among the many outputs from these models. These flows constitute hidden variables have not been observed directly in the archaeological record.

Adopting this perspective on ABMs suggests many new ways they can be combined with stochastic process modeling. Most obviously, if our original dataset has lots of missing values, by fitting the parameters of an ABM to match the data we *do* have, we could use the values the ABM assigns to the variables with missing data as estimates of those missing data values, a technique proposed by Bruch and Atwell (Bruch and Atwell 2015) for social science studies more broadly. In this sense, ABMs can sometimes be used as a variant of the technique of imputations, used so extensively to deal with missing data in the original Seshat analysis (Turchin *et al.* 2018a). As another example, we could sample the values of the hidden variables in the ABM at regular time-intervals. By adding those samples to the original dataset, we could produce an estimated dataset in a much larger space than the original dataset. For example, if we fit the parameters of an ABM so that it reproduced the data in Seshat, then we could sample the variables of the ABM at single-century intervals, and add those sampled values to Seshat, to produce a dataset in a much larger space. We could then perform any of the stochastic process analyses described above to this augmented dataset. For example, we could perform PCA on this augmented dataset, and examine how the PC2 of this new PCA varies as the new PC1 increases, perhaps finding a more elaborate version of the hinge points that were found by analyzing the original Seshat dataset.

5 CONCLUSION

The concept of history unfolding stochastically is not new; in the context of the history of life on Earth, Stephen J. Gould famously asked what would happen if we could "replay the tape", which implicitly supposes that an underlying stochastic process generated that tape (Gould 1989). Similarly, stochastic process modeling of environmental dynamics has been used to infer exogenous perturbations to the dynamics of human social systems (Malik 2020). In addition, the phylogenetic tree reconstructions of human language dynamics discussed in Section IIIB have been used as a "clock" to infer the dynamics of socio-political phenomena (Currie *et al.* 2010, Sheehan *et al.* 2018). There has also been some work directly applying time-series analysis techniques to socio-political datasets (Turchin 2018).

These are isolated instances though, rather than a systematic scientific program. Here we propose something more fundamental: that by grounding our investigations of human social dynamics in stochastic process models, we can not only better investigate the historical record, but also begin to unify the myriad approaches that have been championed for analyzing that record. Such a program would also potentially allow us to detect drivers for the historical processes that generated that historical record — in particular, drivers that had not already been anticipated in social science models. This might allow the data to drive our formulation of social science models, as an adjunct to the more conventional approach under which we analyze datasets only after we first formulate models (e.g., based on intuitive insight and/or on analogizing with models from other scientific fields). Crucially, as we illustrated with the examples above, both the datasets and computational tools necessary for this vision to become a reality are now coming into being.

It is important to emphasize that we do *not* argue that one specific stochastic process we have identified generates the dynamics of history. (Indeed, we expect that it will be most fruitful to view history as multiple, interwoven stochastic processes, all with different characteristics.) We are not even advocating whether a time-homogeneous process or time-inhomogeneous process be considered. Ultimately, as in all statistical analyses, the choice of model to fit to the data is governed by considerations of number of data, size of the space of variables, types of variables, etc., with cross-validation used to help winnow the options. (See supplement, Section IA.)

We are also not advocating that one specific state space be used to model the stochastic process(es) of human history. Nor are we arguing which subsequent analyses should be applied to stochastic process(es) models inferred from historical data, e.g., to uncover possible causal relationships among historical variables.

More generally, we are also *not* arguing that all of historical analysis should be formulated in terms of stochastic processes. Statics, as opposed to dynamics, is also (obviously) an extremely important aspect of historical analysis, as history is not only concerned with time series analysis, but also with revealing the internal structures of societies and the patterns of their interactions at any given single point of time. Indeed, even in those sciences where all phenomena are based on a single dynamic law, like quantum physics (Schrödinger's equation), much research focuses on statics rather than dynamics.

Finally, we note that a stochastic process formulation is also central to the other historical sciences, ranging from biology to meteorology to geology. So not only does this perspective allow us to unify the analyses of computational history, it also allows us to align how we investigate human history with how it is done in the other historical sciences.

DATA ACCESSIBILITY STATEMENT

For the facilitation of open science and the broader research community, we provide transparent access to key resources integral to this article:

1. bighist Python Package: A cornerstone of our analysis pipeline, the 'bighist' Python package, not only offers a data abstraction framework suitable for analyses in big history and cultural evolution but also contains the actual Seshat data necessary to reproduce our results. The package simplifies the process of loading and manipulating the seshat data, ensuring its practical utility for researchers. The source code, alongside comprehensive installation guidelines, is available here.
2. Article Source Code: For hands-on replication and a deeper understanding of our methods and findings, much of our article-specific source code and examples are openly accessible here.
3. Additional information: Please see the supplement for further details, especially the section "Source code and the Python bighist package."

NOTES

1 Of course, our preconceptions and experience still affect our choice of which variables go into the historical dataset in the first place. Our concern here is to able to minimize the role of such preconceptions in the analysis of our datasets, however those datasets were constructed.

2 This need of high-order Markov models to have very large amounts of data is similar to the need of conventional delay embedding time-series analysis for such large datasets. See the discussion section below.

3 To be more concrete, suppose that the stochastic process itself changes at time t'. This would mean that if we start the society at $\mathbf{x}_0$ at some time $t_1 > t'$, the likely ensuing trajectories would be different from what they would be if we instead start the society at $\mathbf{x}_0$ at a time $t_0 < t'$.

4 As a technical comment, throughout this paper we view real-world datasets as being formed by noisy observations of trajectories $\mathbf{x}(t)$, rather than incorporating the observational process directly into the stochastic process itself. So we do not consider the possible effects of time-dependence in the map taking $\mathbf{X}(t)$ to our observational data. In particular, for the purposes of this paper, we do not consider the many ways that archaeological data concerning events further in the past can be "more noisy" than recent archaeological data.

5 Note though that such techniques can also be interpreted as estimating deterministic dynamics embedded on a high dimensional manifold in state-space (Takens 1981; Sauer et al. 1991). At present it is unclear how low-order Markov models in which the drift and diffusion vary across the space are related to such processes generated by deterministic dynamics over a high dimensional manifold.

6 We emphasize, though, that we have not yet done a careful analysis of exactly how much correlation there is.

7 A time-homogeneous Markov chain has a unique stationary probability distribution. That distribution can have multiple peaks, but the probability is 0 of the system staying in one peak indefinitely. In that sense, the dynamics cannot have multiple basins of attraction.

ADDITIONAL FILE

The additional file for this article can be found as follows:

- **Supplement File.** Supplementary Information for The Past as a Stochastic Process. DOI: https://doi.org/10.5334/jcaa.113.s1

ACKNOWLEDGEMENTS

We thank S. DeDeo for useful conversations in the early stages of formulating this paper.

FUNDING INFORMATION

David H. Wolpert: Santa Fe Institute.
Stefani A. Crabtree: Coalition for Archaeological Synthesis.

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

David H. Wolpert orcid.org/0000-0003-3105-2869
Santa Fe Institute, US; Complexity Science Hub Vienna, AT; Arizona State University, US

Michael H. Price
Channel Cognition, LLC, US

Stefani A. Crabtree orcid.org/0000-0001-8585-8943
Utah State University, US; The Santa Fe Institute, US; Australian Research Council Centre of Excellence for Australian Biodiversity and Heritage, Australia; Crow Canyon Archaeological Center, US

Timothy A. Kohler orcid.org/0000-0002-3414-6660
Santa Fe Institute, US; Washington State University, US; Crow Canyon Archaeological Center, US

Jürgen Jost orcid.org/0000-0001-5258-6590
Santa Fe Institute, US; The Max Planck Institute for Mathematics in the Science, DE

James Evans
Santa Fe Institute, US; University of Chicago, US

Peter F. Stadler orcid.org/0000-0002-5016-5191
Santa Fe Institute, US; The Max Planck Institute for Mathematics in the Science, DE; Leipzig University, DE

Hajime Shimao
Santa Fe Institute, US

Manfred D. Laubichler orcid.org/0000-0001-6152-0251
Santa Fe Institute, US; Arizona State University, US

REFERENCES

Aimers, JJ. 2007. What Maya collapse? Terminal classic variation in the Maya lowlands. *Journal of Archaeological Research*, 15: 329–377. DOI: https://doi.org/10.1007/s10814-007-9015-x

Alatas, SF. 2013. Ibn khaldun. DOI: https://doi.org/10.1093/acprof:oso/9780198090458.001.0001

Amblard, P-O and **Michel, OJJ.** 2013. The relation between Granger causality and directed information theory: A review. *Entropy*, 15: 113–143. DOI: https://doi.org/10.3390/e15010113

Arfken, GB and **Weber, HJ.** 1999. Mathematical methods for physicists.

Baćak, V and **Kennedy, EH.** 2019. Principled machine learning using the super learner: an application to predicting prison violence. *Sociological Methods & Research*, 48: 698–721. DOI: https://doi.org/10.1177/0049124117747301

Bainbridge, WS. 1995. Neural network models of religious belief. *Sociological perspectives*, 38: 483–495. DOI: https://doi.org/10.2307/1389269

Barron, ATJ, Huang, J, Spang, RL and **DeDeo, S.** 2018. Individuals, institutions, and innovation in the debates of the French Revolution. *Proceedings of the National Academy of Sciences*, 115: 4607–4612. DOI: https://doi.org/10.1073/pnas.1717729115

Beheim, B, Atkinson, QD, Bulbulia, J, Gervais, W, Gray, RD, Henrich, J, Lang, M, Monroe, MW, Muthukrishna, M, Norenzayan, A, et al. 2021. Treatment of missing data determined conclusions regarding moralizing gods. *Nature*, 595: E29–E34. DOI: https://doi.org/10.1038/s41586-021-03655-4

Bhattacharya, T, Blasi, D, Croft, W, Cysouw, M, Hruschka, D, Maddieson, I, Müller, L, Retzlaff, N, Smith, E, Stadler, PF, Starostin, G and **Youn, H.** 2018. Studying Language Evolution in the Age of Big Data. *J. Language Evol.*, 3: 94–129. DOI: https://doi.org/10.1093/jole/lzy004

Blast, DE, Moran, S, Moisik, SR, Widmer, P, Dediu, D and **Bickell, B.** 2019. Human sound systems are shaped by post-Neolithic changes in bite configuration. *Science*, 363: eaav3218. DOI: https://doi.org/10.1126/science.aav3218

Bruch, E and **Atwell, J.** 2015. Agent-based models in empirical social research. *Sociological Methods & Research*, 44: 186–221. DOI: https://doi.org/10.1177/0049124113506405

Carley, KM. 1996. Artificial intelligence within sociology. *Sociological methods & research*, 25: 3–30. DOI: https://doi.org/10.1177/0049124196025001001

Caswell, H. 2001. *Matrix Population Models: Construction, Analysis and Interpretation*, 2nd ed. Sunderland, MA: Sinauer.

Crabtree, SA, Bird, DW and **Bird, RB.** 2019. Subsistence Transitions and the Simplification of Ecological Networks in the Western Desert of Australia. *Human Ecology*, 47: 165–177. DOI: https://doi.org/10.1007/s10745-019-0053-z

Crabtree, SA, Bocinsky, RK, Hooper, PL, Ryan, SC and **Kohler, TA.** 2017b. How to Make a Polity (in the central Mesa Verde Region). *American Antiquity*, 82: 71–95. DOI: https://doi.org/10.1017/aaq.2016.18

Crabtree, SA, Vaughn, LJS and **Crabtree, NT.** 2017a. Reconstructing Ancestral Pueblo food webs in the southwestern United States. *Journal of Archaeological Science*, 81: 116–127. DOI: https://doi.org/10.1016/j.jas.2017.03.005

Craigmile, P, Herbei, R, Liu, G and **Schneider, G.** 2022. Statistical inference for stochastic differential equations. *Wiley Interdisciplinary Reviews: Computational Statistics*, e1585. DOI: https://doi.org/10.1002/wics.1585

Currie, TE, Greenhill, SJ, Gray, RD, Hasegawa, T and **Mace, R.** 2010. Rise and fall of political complexity in island south-east asia and the pacific. *Nature*, 467: 801–804. DOI: https://doi.org/10.1038/nature09461

Davidson, T. 2017. Black Box Models and Sociological Explanations: Predicting GPA Using Neural Networks. DOI: https://doi.org/10.31235/osf.io/7nsrf

Davis, W and **Buffett, B.** 2022. Estimation of drift and diffusion functions from unevenly sampled time-series data. *Physical Review E*, 106: 014140. DOI: https://doi.org/10.1103/PhysRevE.106.014140

Dunne, JA, Maschner, H, Betts, MW, Huntly, N, Russell, R, Williams, RJ and **Wood, SA.** 2016. The roles and impacts of human huntergatherers in North Pacific marine food webs. *Scientific Reports*, 6: 21179. DOI: https://doi.org/10.1038/srep21179

Durham, WH. 1991. *Coevolution: Genes, Culture, and Human Diversity*. Stanford University Press. DOI: https://doi.org/10.1515/9781503621534

Francois, P, Manning, JG, Whitehouse, H, Brennan, R, Currie, T, Feeney, K and **Turchin, P.** 2016. A macroscope for global history: Seshat global history databank, A methodological overview.

Friedrich, R, Peinke, J, Sahimi, M and **Tabar, MRR.** 2011. Approaching complexity by stochastic methods: From biological systems to turbulence. *Physics Reports*, 506: 87–162. DOI: https://doi.org/10.1016/j.physrep.2011.05.003

Frishman, A and **Ronceray, P.** 2020. Learning force fields from stochastic trajectories. *Physical Review X*, 10: 021009. DOI: https://doi.org/10.1103/PhysRevX.10.021009

Geertz, C. 1973. Thick description: Toward and interpretive theory of culture. In: Geertz, C (ed.), *The Interpretation of Cultures*, Chap. 1. Basic Books. pp. 3–30.

Gould, SJ. 1989. *Wonderful Life: The Burgess Shale and the Nature of History*. New York: W. W. Norton.

Granger, CW. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37: 424–438. DOI: https://doi.org/10.2307/1912791

Gray, RD, Atkinson, QD and **Greenhill, SJ.** 2011. Language evolution and human history: what a difference a date makes. *Phil. Trans. Roy. Soc. B*, 366: 1090–1100. DOI: https://doi.org/10.1098/rstb.2010.0378

Harper, K. 2021. *Plagues Upon the Earth: Disease and the Course of Human History*. Princeton University Press. DOI: https://doi.org/10.1515/9780691224725

Hodder, I. 2012. *Entangled: An Archaeology of the Relationships between Humans and Things*. Chichester, UK: Wiley-Blackwell. DOI: https://doi.org/10.1002/9781118241912

Hoffmann, M, Scherer, M, Hempel, T, Mardt, A, de Silva, B, Husic, BE, Klus, S, Wu, H, Kutz, N, Brunton, SL, et al. 2021. Deeptime: a python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology*, 3: 015009. DOI: https://doi.org/10.1088/2632-2153/ac3de0

Hoggarth, JA, Restall, M, Wood, JJ and **Kennett, DJ.** 2017. Drought and its demographic effects in the maya lowlands. *Current Anthropology*, 58: 82–113. DOI: https://doi.org/10.1086/690046

Ingold, T. 2007. *Lines: A Brief History*. New York: Routledge. DOI: https://doi.org/10.4324/9780203961155

Jones, JH. 2009. The force of selection on the human life cycle. *Evolution and Human Behavior*, 30: 305–314. DOI: https://doi.org/10.1016/j.evolhumbehav.2009.01.005

Jones, JH and **Tuljapurkar, S.** 2015. Measuring selective constraint on fertility in human life histories. *Proc. Natl. Acad. Sci.,* 112: 8982–8986. USA. DOI: https://doi.org/10.1073/pnas.1422037112

Kennett, DJ, Breitenbach, SFM, Aquino, VV, Asmerom, Y, Awe, J, Baldini, JUL, Bartlein, P, Culleton, BJ, Ebert, C, Jazwa, C, et al. 2012. Development and disintegration of Maya political systems in response to climate change. *Science*, 338: 788–791. DOI: https://doi.org/10.1126/science.1226299

Khaldun, I. 2015. *The Muqaddimah: An Introduction to History-Abridged Edition*. Princeton University Press.

Kidger, P, Foster, J, Li, X, Oberhauser, H and **Lyons, T.** 2020. Neural sdes made easy: Sdes are infinite-dimensional gans. https://openreview.net/forum?id=padYzanQNbg.

Kohler, TA, Crabtree, SA, Bocinsky, RK and **Hooper, PL.** 2018. Sociopolitical Evolution in Mid-Range Societies: The Prehispanic Pueblo Case. In: Sabloff, JA and Sabloff, PLW (eds.), *The Emergence of Premodern States: New Perspectives on the Development of Complex Societies*. Santa Fe, NM: Santa Fe Institute. pp. 133–184. DOI: https://doi.org/10.37911/9781947864030.06

Lawler, GF. 2018. *Introduction to stochastic processes*. CRC Press. DOI: https://doi.org/10.1201/9781315273600

Li, X, Wong, T-KL, Chen, RT and **Duvenaud, DK.** 2020. Scalable gradients and variational inference for stochastic differential equations. In: *Symposium on Advances in Approximate Bayesian Inference*, PMLR. pp. 1–28.

Lütkepohl, H. 2005. *New Introduction to Multiple Time Series Analysis*. Springer Science & Business Media. DOI: https://doi.org/10.1007/978-3-540-27752-1

MacLeod, H, Yang, S, Oakes, K, Connelly, K and **Natarajan, S.** 2016. Identifying rare diseases from behavioural data: a machine learning approach. In: *2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE)*, IEEE, pp. 130–139. DOI: https://doi.org/10.1109/CHASE.2016.7

Malik, N. 2020. Uncovering transitions in paleoclimate time series and the climate driven demise of an ancient civilization. *Chaos: An Interdisciplinary Journal of Nonlinear Science,* 30: 083108. DOI: https://doi.org/10.1063/5.0012059

Maltseva, AV, Shilkina, NE and **Mahnitkina, OV.** 2016. Data mining sociology: experience and outlook for research. *Sociological Studies*, 3: 35–44.

McAllister, JW. 2002. Historical and Structural Approaches in the Natural and Social Sciences. In: Tindemans, P, Verrijn-Stuart, A and Visser, R (eds.), *The Future of the Sciences and Humanities*. Amsterdam: Amsterdam University Press. pp. 19–54.

Meyn, S and **Tweedie, RL.** 2009. *Markov Chains and Stochastic Stability*. Cambridge University Press. pp. 4868–4879. DOI: https://doi.org/10.1017/CBO9780511626630

Miranda, L and **Freeman, J.** 2020. The two types of society: Computationally revealing recurrent social formations and their evolutionary trajectories. *PLOS ONE,* 15: e0232609. DOI: https://doi.org/10.1371/journal.pone.0232609

Painter, DT, Kempes, C, West, G and **Laubichler, M.** 2020. The four regimes of the Great Acceleration. *Anthropocence Review*. In review.

Pavliotis, GA. 2016. *Stochastic processes and applications*. Springer.

Peregrine, PN. 2018. Toward a Theory of Recurrent Social Formations. In: Sabloff, JA and Sabloff, PLW (eds.), *The Emergence of Pre-Modern States: New Perspectives on the Development of Complex Societies*. Santa Fe, NM: SFI Press. pp. 271–295. DOI: https://doi.org/10.37911/9781947864030.09

Peregrine, PN. 2020. Climate and social change at the start of the Late Antique Little Ice Age. DOI: https://doi.org/10.1177/0959683620941079

Purzycki, BG, Apicella, C, Atkinson, QD, Cohen, E, McNamara, RA, Willard, AK, Xygalatas, D, Norenzayan, A and **Henrich, J.** 2016. Moralistic gods, supernatural punishment and the expansion of human sociality. *Nature*, 530: 327–330. DOI: https://doi.org/10.1038/nature16980

Rogers, A. 1966. The Multiregional Matrix Growth Operator and the Stable Interregional Age Structure. *Demography*, 3: 537–544. DOI: https://doi.org/10.2307/2060178

Romanowska, I, Gamble, C, Bullock, S and **Sturt, F.** 2017. Dispersal and the movius line: Testing the effect of dispersal on population density through simulation. *Quaternary International*, 431(B): 53–63. DOI: https://doi.org/10.1016/j.quaint.2016.01.016

Sauer, T, Yorke, JA and **Casdagli, M.** 1991. Embedology. *Journal of statistical Physics*, 65: 579–616. DOI: https://doi.org/10.1007/BF01053745

Scheinfeldt, LB, Soi, S and **Tishkoff, SA.** 2010. Working toward a synthesis of archaeological, linguistic, and genetic data for inferring African population history. *Proc. Natl. Acad. Sci*., 107(S2): 8931–8938. USA. DOI: https://doi.org/10.1073/pnas.1002563107

Schreiber, T. 2000. Measuring information transfer. *Physical Review Letters*, 85: 461–464. DOI: https://doi.org/10.1103/PhysRevLett.85.461

Sewell, WH. 1992. A Theory of Structure: Duality, Agency, and Transformation. *American Journal of Sociology*, 98: 1–29. DOI: https://doi.org/10.1086/229967

Sheehan, O, Watts, J, Gray, RD and **Atkinson, QD.** 2018. Coevolution of landesque capital intensive agriculture and sociopolitical hierarchy. *Proceedings of the National Academy of Sciences*, 115: 3628–3633. DOI: https://doi.org/10.1073/pnas.1714558115

Shin, J, Price, MH, Wolpert, DH, Shimao, H, Tracey, B and **Kohler, TA.** 2020. Scale and information-processing thresholds in Holocene social evolution. *Nat. Comm*. 11: 2394. DOI: https://doi.org/10.1038/s41467-020-16035-9

Steffen, W, Sanderson, A, Tyson, PD, Jager, J, Matson, PA, III Moore, B, Oldfield, F, Richardson, K, Schellnhuber, HJ, Turner, BL and **Wasson, RJ.** 2004. *Global Change and the Earth System: A Planet Under Pressure*. Springer-Verlag, Berlin. DOI: https://doi.org/10.1007/b137870

Strawhacker, C, Snitker, G, Peeples, M, Kinzig, A, Kintigh, K, Bocinsky, K and **Spielmann, K.** 2020. A Landscape Perspective on Climate-Driven Risks to Food Security: Exploring the Relationship between Climate and Social Transformation in the Prehispanic U.S. Southwest. *American Antiquity*, 85: 427–451. DOI: https://doi.org/10.1017/aaq.2020.35

Tainter, JA. 1988. *The Collapse of Complex Societies*, Vol. New Studies in Archaeology. Cambridge University Press.

Takens, F. 1981. Detecting strange attractors in turbulence. In: *Dynamical systems and turbulence, Warwick 1980*. Springer. pp. 366–381. DOI: https://doi.org/10.1007/BFb0091924

Turchin, P, Currie, TE, Whitehouse, H, Francois, P, Feeney, K, Mullins, D, Hoyer, D, Collins, C, Grohmann, S, Savage, P, Mendel-Gleason, G, Turner, E, Dupeyron, A, Cioni, E, Reddish, J, Levine, J, Jordan, G, Brandl, E, Williams, A, Cesaretti, R, Krueger, M, Ceccarelli, A, Figliulo-Rosswurm, J, Tuan, P-J, Peregrine, P, Marciniak, A, Preiser-Kapeller, J, Kradin, N, Korotayev, A, Palmisano, A, Baker, D, Bidmead, J, Bol, P, Christian, D, Cook, C, Covey, A, Feinman, G, Júlíusson, AD, Kristinsson, A, Miksic, J, Mostern, R, Petrie, C, Rudiak-Gould, P, ter Haar, B, Wallace, V, Mair, V, Xie, L, Baines, J, Bridges, E, Manning, J, Lockhart, B, Bogaard, A and **Spencer, C.** 2018b. Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization - supplementary material. *Proc. Natl. Acad. Sci*. 115: E144–E151. Supplementary Material. USA. DOI: https://doi.org/10.1073/pnas.1708800115

Turchin, P. 2018. Fitting dynamic regression models to seshat data. *Cliodynamics*, 9. DOI: https://doi.org/10.21237/C7CLIO9137696

Turchin, P, Hoyer, D, Bennett, J, Basava, K, Cioni, E, Feeney, K, Francois, P, Holder, S, Levine, J, Nugent, S, et al. 2020. The equinox2020 seshat data release. *Cliodynamics,* 11. DOI: https://doi.org/10.21237/C7CLIO11148620

Turchin, P, Currie, TE, Whitehouse, H, Francois, P, Feeney, K, Mullins, D, Hoyer, D, Collins, C, Grohmann, S, Savage, P, Mendel-Gleason, G., Turner, E, Dupeyron, A, Cioni, E, Reddish, J, Levine, J, Jordan, G, Brandl, E, Williams, A, Cesaretti, R, Krueger, M, Ceccarelli, A, Figliulo-Rosswurm, J, Tuan, P-J, Peregrine, P, Marciniak, A, Preiser-Kapeller, J, Kradin, N, Korotayev, A, Palmisano, A, Baker, D, Bidmead, J, Bol, P, Christian, D, Cook, C, Covey, A, Feinman, G, Júlíusson, AD, Kristinsson, A, Miksic, J, Mostern, R, Petrie, C, Rudiak-Gould, P, ter Haar, B, Wallace, V, Mair, V, Xie, L, Baines, J, Bridges, E, Manning, J, Lockhart, B, Bogaard, A and **Spencer, C.** 2018a. Quantitative historical analysis uncovers a single dimension of complexity that structures global variation in human social organization. *Proc. Natl. Acad. Sci*., 115: E144–E151. USA. DOI: https://doi.org/10.1073/pnas.1708800115

Turner, BL and **Sabloff, JA.** 2012. Classic period collapse of the central maya lowlands: Insights about human–environment relationships for sustainability. *Proceedings of the National Academy of Sciences*, 109: 13908–13914. DOI: https://doi.org/10.1073/pnas.1210106109

Tylor, EB. 1871. *Primitive Culture: Researches into the Development of Mythology, Philosophy, Religion, Art, and Custom*. London: John Murra.

Webster, D. 2002. *The Fall of the Ancient Maya: Solving the Mystery of the Maya Collapse*. Thames & Hudson.

Weiss, H, Courty, MA, Wetterstrom, W, Guichard, F, Senior, L, Meadow, R and **Curnow, A.** 1993. The Genesis and Collapse of Third Millennium North Mesopotamian Civilization. *Science*, 261: 995–1004. DOI: https://doi.org/10.1126/science.261.5124.995

Whitehouse, H, Francois, P, Savage, PE, Currie, TE, Feeney, KC, Cioni, E, Purcell, R, Ross, RM, Larson, J, Baines, J, et al. 2019. Complex societies precede moralizing gods throughout world history. *Nature*, 568: 226–229. DOI: https://doi.org/10.1038/s41586-019-1043-4

Whitehouse, H, Francois, P, Savage, PE, Currie, TE, Feeney, KC, Cioni, E, Purcell, R, Ross, RM, Larson, J, Baines, J, et al. 2021. Retraction note: Complex societies precede moralizing gods throughout world history. *Nature*, 595: 320–320. DOI: https://doi.org/10.1038/s41586-021-03656-3

Wood, JW. 1998. A theory of preindustrial population dynamics demography, economy, and well-being in Malthusian systems. *Current Anthropology*, 39: 99–135. DOI: https://doi.org/10.1086/204700

Yeakel, JD, Pires, MM, Rudolf, L, Dominy, NJ, Koch, PL, Guimarães, PR and **Gross, T.** 2014. Collapse of an ecological network in Ancient Egypt. *Proceedings of the National Academy of Sciences*, 111: 14472–14477. https://www.pnas.org/content/111/40/14472.full.pdf. DOI: https://doi.org/10.1073/pnas.1408471111

Yildiz, C, Heinonen, M, Intosalmi, J, Mannerstrom, H and **Lahdesmaki, H.** 2018. Learning stochastic differential equations with Gaussian processes without gradient matching. In: *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP),* IEEE. pp. 1–6. DOI: https://doi.org/10.1109/MLSP.2018.8516991

Zeng, L. 1999. Prediction and classification with neural network models. *Sociological methods & research*, 27: 499–524. DOI: https://doi.org/10.1177/0049124199027004002

Zhan, D, Yi, S and **Jiang, D.** 2018. Small-Scale Demographic Sequences Projection Based on Time Series Clustering and LSTM-RNN. In: *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE. pp. 803–809. DOI: https://doi.org/10.1109/ICDMW.2018.00120

TO CITE THIS ARTICLE:
Wolpert, DH, Price, MH, Crabtree, SA, Kohler, TA, Jost, J, Evans, J, Stadler, PF, Shimao, H and Laubichler, MD. 2024. The Past as a Stochastic Process. *Journal of Computer Applications in Archaeology,* 7(1): 134–152. DOI: https://doi.org/10.5334/jcaa.113

Submitted: 06 May 2023 **Accepted:** 21 November 2023 **Published:** 08 February 2024