

Perspective

A deeper look into natural sciences with physics-based and data-driven measures

Davi Röhe Rodrigues,¹ Karin Everschor-Sitte,^{1,*} Susanne Gerber,² and Illia Horenko³

SUMMARY

With the development of machine learning in recent years, it is possible to glean much more information from an experimental data set to study matter. In this perspective, we discuss some state-of-the-art data-driven tools to analyze latent effects in data and explain their applicability in natural science, focusing on two recently introduced, physics-motivated computationally cheap tools—latent entropy and latent dimension. We exemplify their capabilities by applying them on several examples in the natural sciences and show that they reveal so far unobserved features such as, for example, a gradient in a magnetic measurement and a latent network of glymphatic channels from the mouse brain microscopy data. What sets these techniques apart is the relaxation of restrictive assumptions typical of many machine learning models and instead incorporating aspects that best fit the dynamical systems at hand.

INTRODUCTION

Being able to explore further length scales than those accessible for the human eye (limited to the range of millimeters to kilometers) allowed for a better understanding of natural sciences. Experimental techniques have evolved significantly to explore the different space scales, from new generation electron microscopes to deep-field telescopes. However, a great challenge imposed by increasing experimental instruments' sensitivity is the excess of information that must be filtered to study the desired behavior. For this reason, there is great demand for a simultaneous development of computational machine learning (ML) techniques to analyze data. The understanding of patterns hidden in the noise has been awarded at least two Physics Nobel Prizes—the discovery of cosmic microwave background (Penzias and Wilson, 1965) in 1978 and gravitational waves (Abbott et al., 2017) in 2017—and lead, for example, to the first image of a black hole (The Event Horizon Telescope Collaboration, 2019a) in 2019. These recent successes rely enormously on the progress achieved in data inference and ML techniques.

While ML has progressed substantially within the last decades, being used across all fields of natural sciences and becoming part of our daily life, the field still faces several challenges. We emphasize two challenges: (i) the computational costs, which still go hand in hand with the ever-increasing size of the data sets, and (ii) often assumed properties restricting the data analysis, such as independency and identically distribution of the data (Blum and Langley, 1997; Al-Jarrah et al., 2015). Thus, there is an increasing demand for high-performance data inference tools that are both computationally cheap and not overly restrictive. Another main issue of several ML methods, such as deep learning, is the lack of inherent understanding of the rules underlying the learning during the data inference process (Rudin, 2019; Papernot et al., 2017). This is an obstacle for controlled and human-understandable improvement of the emerging methods (Deng et al., 2020). ML has strong links to information theory (such as the Shannon entropy, (log-) likelihood, and information content), which borrows many concepts of theoretical physics, particularly thermodynamics. The main goal of computing tools for data extraction is to (i) remove irrelevant data, usually associated with noise, and (ii) recognize patterns and behaviors in the underlying relevant data. These two processes can be carried out in a multitude of ways with tools that diverge, as follows, on their way to analyze and interpret the data. Concerning the first process, there are denoising tools (Buades et al., 2005; Shao et al., 2013; Goyal et al., 2020). For the second process, typically one assumes a model for the data such that one can infer latent behaviors. Some models are very powerful and broadly used such as Gaussian mixture models (GMMs) or hidden Markovian models (HMMs). However, imposing a model implies assumptions on the data and constrains information extraction. The most frequent assumptions are the identical and independent distribution (i.i.d.), as well as, Gaussianity of the data. Recently,

¹Institute of Physics, Johannes Gutenberg University of Mainz, 55128 Mainz, Germany

²Institute of Human Genetics, University Medical Center of the Johannes Gutenberg University Mainz, 55131 Mainz, Germany

³Università della Svizzera Italiana, Faculty of Informatics, Via G. Buffi 13, 6900 Lugano, Switzerland

*Correspondence: kaeversc@uni-mainz.de
<https://doi.org/10.1016/j.isci.2021.102171>



several physics-inspired models have been proposed in order to take into account assumptions that best fit the observed dynamical systems (Williams et al., 2015; Runge et al., 2015; Ye et al., 2015; Vesselinov et al., 2019; Zenil et al., 2019; Rupe et al., 2019; Horenko et al., 2019). Furthermore, there has been a considerable effort in developing computationally scalable data-driven methods that allow overcoming bias imposed by the restrictive model assumptions of common tools (Gerber et al., 2020; Horenko, 2020). In particular, two novel physics-inspired methods—called latent entropy and latent dimension (Horenko et al., 2019)—not only combine both the denoising and pattern recognition step but also provide a computationally cheap measure for the data that allows for sharp recognition of patterns.

In this perspective, we discuss some state-of-the-art data-driven inference tools for analyzing latent effects in data and explain their applicability in natural science. In particular, in the [Latent methods](#) section, we will introduce (A) denoising tools, (B) model-based data analysis tools, such as GMMs, HMMs, and other physics-inspired models, and (C) two recently introduced, physics-motivated computationally cheap tools—the latent entropy and the latent dimension.

To exemplify the methods, in particular the latent entropy and latent dimension, in the [Applications](#) section, we first demonstrate them on a toy model clarifying the role of the latent entropy as a measure for stochasticity of the underlying dynamic process and the role of the latent dimension as a measure for the system's memory. Then, we apply them to (A) the data-driven visualization of the glymphatic system of a mouse brain from the in vivo brain microscopy data, revealing latent bulk capillaries; (B) an analysis of magneto-optical Kerr effect (MOKE) data from magnetization experiments, revealing a hidden gradient across the sample not visible to common measures; and (C) the analysis of two noisy astrophysical videos revealing features of the M57 (Ring Nebula) in Lira and the M31 star cluster in Hercules.

Latent methods

Experimental measurements are typically performed by probing a sample via a perturbation and observing its response, which e.g., in microscopy amounts to shining light onto the sample and measuring how the sample reflects the light. Therefore, the quantity of interest can often only be captured indirectly through, for example, changes in the incident ray, photon counts, or event rates. The challenge is then to extract the desired information from the data acquired, which are typically affected by other physical processes as well as by noise. There is a major effort in finding the most appropriate methods to access and extract these latent—or hidden—phenomena of the observed system. The latent methods of modern information theory typically examine the statistical properties of the data (Bayes, 1763; Shannon, 1948; Tanner, 2012; Enßlin et al., 2009), inspired by the concepts like entropy—originally coming from statistical physics. In this perspective, we focus on the problem of spatial pattern recognition in video data.

Denoising tools

Denoising tools remove unwanted identifiable noise thereby, allowing for a clearer comprehension of the relevant data (Buades et al., 2005; Shao et al., 2013; Goyal et al., 2020). The main challenge is to eliminate noise without compromising the relevant data. The methods vary greatly from local and non-local filters based on the correlation between pixels (Wiener, 1950; Brailean et al., 1995; Stockman and Shapiro, 2001; Buades et al., 2005; Shao et al., 2008) to mapping the data to other domains where patterns can be recognized (Dabov et al., 2007; Mairal et al., 2009; Luisier et al., 2010). To identify the noise, denoising tools typically require a model for the noise. With this purpose, they assume, for example, an oscillatory behavior or Gaussianity of the noise (Ernst, 1966; Wink and Roerdink, 2004; Luisier et al., 2011; Zhang et al., 2017). Some of the recent approaches attempt to learn a model for the noise based on ML tools (Hinton, 2006). In general, once the features of the data are classified within the model for the noise, the filter removes information that is considered unwanted, such as features with high frequency or that have low probability.

Denoising has become such a common instrument that it has been integrated in most mathematical software. In the section concerning the data analysis of the Glymphatic system of a mouse brain we have, for example, we use the denoising autoencoder of MATLAB (Zhang et al., 2017; Berghout et al., 2020, 2021). In particular, this modified denoise autoencoder (Bengio and LeCun, 2007) is based on a supervised ML algorithm which models noise based on a training set. A filter is then used to remove the features from the data that correspond to the modeled noise.

Even though denoising tools can enhance the visualization of some data features by removing irrelevant information, they are unable to identify latent behaviors in the data.

Model-based data analysis tools

Several model-based data analysis tools have been inspired by information theory and physics principles in order to extract relevant features from the data. Ideally, latent methods not only detect patterns in data, i.e., accurately group together pixels in video data with similar properties, but also are cost effective in terms of computational power. Below we present a few examples.

Gaussian mixture models

GMMs have been successfully implemented in a vast body of research fields, including economics (Barn-dorff-Nielsen and Shephard, 2001; Phillips and Loretan, 1991), genetics (François and Durand, 2010; Loh et al., 2015; Turelli and Barton, 1994), psychology (Bauer, 2007; Hipp and Bauer, 2006; Rachuri et al., 2010), and speech recognition (Torres-Carrasquillo et al., 2002; Reynolds, 1995). They assume that the underlying effects represented by the data are generated from Gaussian distributions characterized by a certain parameter set (Pedregosa et al., 2011; Melnykov and Maitra, 2010). The mean, covariance, and prior probability associated with each Gaussian distribution manifest the presence of latent variables. These parameters can be calculated by different variants of the expectation maximization algorithm (Frühwirth-Schnatter, 2006; Greggio et al., 2012; Pinto and Engel, 2015) which fits the mixture of Gaussian models that best identifies clusters in the analyzed data.

An important characteristic of GMMs is the invariance under reordering or permutation of the data sequence. The key assumptions of GMMs are the independency and identical distribution of the random variables (i.i.d. assumption), as well as homogeneity of the input data. The first two assumptions allow to fit properties of the data by a mix of Gaussian distributions, while the third assumption allows for fitting the data of all the frames with the same number of Gaussian distributions. These properties allow data to be analyzed for each frame independently and to be fitted by a mix of Gaussian distributions. GMMs are thus useful to track objects even when they are absent or occluded in some frames (McKenna et al., 1998; Kaew-TraKulPong & Bowden, 2002; Stauffer and Grimson, 1999; Santosh et al., 2013).

Below in the data analysis of images of the [Glymphatic system of a mouse brain](#), we show how the GMM entropy allows us to recognize surface lymph capillaries from microscopy data of a mouse brain.

Hidden Markov models

HMMs take into account the time sequence of the frames to extract latent—"hidden"—information. They have been successfully applied to economics (Hassan and Nath, 2005; Dias et al., 2015), genetics (Collier et al., 2000; Stanke and Waack, 2003; Xi et al., 2010; Narasimhan et al., 2016), cancer diagnosis (Manogaran et al., 2018), speech recognition (Schuller et al., 2003), and weather research (Bellone et al., 2000). Their restrictive assumption is that the system is assumed to be a time-homogeneous Markov model such that each data frame depends solely on the previous via a stochastic process. More specifically, HMMs assume that the sequence of observables is obtained from a Markov process involving hidden states. By calculating the transition probability between the hidden states and the state observation likelihood from the hidden states, one obtains a model that is capable to predict outputs from given observables (Jurafsky and Martin, 2009). Physical systems, however, often present a more complex dynamics that cannot be captured by the time-homogeneous Markov model assumption.

Physics-inspired models

Physical principles can be leveraged to extract relevant information from data. By incorporating expected physical behaviors as basic assumptions into developing ML tools, it is possible to obtain predictive models for certain physical systems (Williams et al., 2015; Runge et al., 2015; Ye et al., 2015; Vesselinov et al., 2019; Zenil et al., 2019; Rupe et al., 2019). These models analyze the time evolution of the data features to obtain powerful predictive models that take into account subtle hidden behaviors. So far, however, most methods remain as conceptual contributions and algorithmic frameworks (Williams et al., 2015; Vesselinov et al., 2019; Zenil et al., 2019; Rupe et al., 2019). Implementations have been limited to certain physical systems, such as weather science (Runge et al., 2015; Ye et al., 2015; Rupe et al., 2019).

Representation of latent effects in transition probabilities

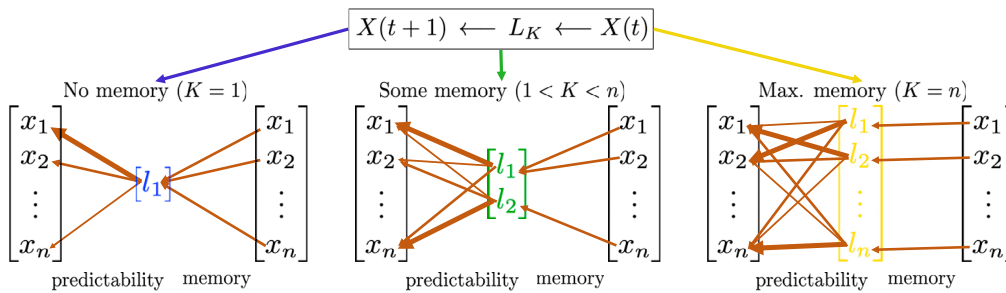


Figure 1. A representation of the meaning of the latent tools

The transition to $X(t+1)$ from $X(t)$ may depend on hidden variables L_K with latent dimension K . The dependence between $X(t+1)$ and $X(t)$, i.e., memory of the system, is related to the dimension K . For example, if $K = 1$, the initial state $X(t)$ is irrelevant and the system has no memory. The distribution of probabilities $\mathbb{P}[X(t+1) = x_j | L_K = l_k]$, represented by the thickness of the arrows, contains information about the predictability of the system.

Scalable probabilistic approximation

A recent proposal toward a realistic implementation without suffering from overly restrictive assumptions from the models is the scalable probabilistic approximation (SPA) algorithms for complex systems (Gerber et al., 2020; Horenko, 2020). SPA algorithms are able to jointly solve clustering, feature selection, and Bayesian model inference problems for data analysis—providing discrete regular clustering approximations that are optimal for Bayesian prediction and classification, with a linear scaling of computational cost and with a parallel communication cost proven to be independent of data size (Gerber et al., 2020; Horenko, 2020). The mathematical properties of the obtained optimal SPA solutions (like regularity) have been demonstrated. The suboptimality of the clustering discretizations obtained with common methods like HMMs when compared to SPA, both in terms of quality and cost, has been proven mathematically. An important element to successfully implement a model is the ability to classify the data features according to a quantitative parameter, i.e., a measure. A measure can be used to cluster the data as well as to obtain insights into the observed system. For this reason, in the following, we focus on two recently introduced measures.

Data measures: latent entropy and latent dimension

The latent entropy and latent dimensions are physics-inspired ML tools for reliable pattern recognition in video data (Horenko et al., 2019). Their algorithms incorporate thermodynamical principles to encode latent properties of the underlying system into two quantitative measures. They have been successfully implemented to analyze several systems, with different levels of complexity and spatiotemporal scales. It has been demonstrated that these tools detect even subtle patterns in data, such as revealing differences of up to 1% in material parameter change in a two-dimensional Ising model even far above the critical temperature (Horenko et al., 2019). They remain accurate also for large noise to signal ratios and are computationally cheap.

The two latent tools operate beyond the sometimes restrictive i.i.d. and Gaussianity assumptions, while taking into account the temporal ordering of the data to infer possible imprints of latent dynamic processes. Unlike GMMs which cluster data points by identifying common properties within each frame, the latent entropy and latent dimension take into account the dynamical evolution between frames to identify patterns in the data. The latter accesses more directly dynamical patterns such as density flows (see section on the [Glymphatic system of a mouse brain](#)) and change in the underlying physical parameters (see section on [temperature gradient in magnetization experiment](#)). The temporal ordering can be examined in terms of memory and predictability, i.e., the correlation between an initial and final state and the tendency of a system to remain in a set of states given the initial conditions, see Figure 1. The system's memory is characterized by the latent dimension while the latent entropy encodes the system's predictability. The sharp detection of latent processes for each pixel then allows for an accurate identification of patterns in the video data when considering the full spatial image.

We consider that each pixel at a time instance t can be described by $X(t)$, an n -dimensional Euclidean vector. The components of $X(t)$ are given by the probabilities $\mathbb{P}[X(t) = x_j]$ of the pixel to assume a value x_j among n categories $\{x_1, x_2, \dots, x_n\}$. These n categories correspond to the discretized color scale of the pixel

value. The time evolution of each pixel in time is, thus, given as a categorical sequence $X = \{X(t=1), X(t=2), \dots, X(t=N)\}$. The probability for X to attain a certain category at a certain time is given by the “exact law of the total probability” (Gardiner, 2004):

$$\mathbb{P}[X(t+1) = x_j] = \sum_{i=1}^n (\Lambda_K)_{ji} \mathbb{P}[X(t) = x_i], \quad (\text{Equation 1})$$

where $(\Lambda_K)_{ji}$ is the transition matrix describing the transition from the initial pixel state to its consecutive one through a potential latent process L_K

$$(\Lambda_K)_{ji} \equiv \sum_{k=1}^K \mathbb{P}[X(t+1) = x_j | L_K = l_k] \mathbb{P}[L_K = l_k | X(t) = x_i]. \quad (\text{Equation 2})$$

Here, $\mathbb{P}[Z = z_i | W = w_j]$ denotes the conditional probability of Z assuming z_i while W is assuming w_j . The latent process $L_K(t)$ takes values from K latent categories $l = \{l_1, l_2, \dots, l_K\}$ where K can assume any value between $(1, \dots, n)$ (Hofmann, 1999, 2001), see Figure 1. In the following, we will use the shorthand notation $(\lambda_K)_{jk} = \mathbb{P}[X(t+1) = x_j | L_K = l_k]$ and $(\gamma_K)_{ki} = \mathbb{P}[L_K = l_k | X(t) = x_i]$ such that $\Lambda_K = \lambda_K \gamma_K$. Given this setting, the algorithm to compute the latent entropy and latent dimension as described in the study by (Horenko et al., 2019) is then given as follows (The MATLAB code is available on https://www.dropbox.com/s/w3few6elo9soegz/MATLAB_Code.zip?dl=0).

- Step 1: Compute the transition matrices Λ_K based on the direct Bayesian model reduction (Hofmann, 1999, 2001; Ding et al., 2006; Gerber & Horenko, 2015, 2017; Gerber et al., 2018), as well as the quantities $S_K = -\frac{1}{N} \log \Lambda_K$ for every K going from 1 to n (To be precise, $S_K = -\sum_{i=1}^n \sum_{j=1}^n C_{ij} \log[(\lambda_K \gamma_K)_{ij}]$, where $C_{ij} = \frac{1}{N} \sum_{t=1}^N \chi(Y(t) = y_i) \chi(X(t) = x_j)$ with χ being an indicator function, is the average contingency table of the data X and Y).
- Step 2: Determine the posterior probabilities p_K for the different latent dimensions $K = 1, \dots, n$ by means of the Akaike information criterion (Hurvich and Tsai, 1989) as follows:

$$p_K = \frac{\exp(-(\text{AIC}_{C_K} - \min_K \text{AIC}_{C_K}))}{\sum_{K=1}^n \exp(-(\text{AIC}_{C_K} - \min_K \text{AIC}_{C_K}))}, \quad (\text{Equation 3})$$

where $\text{AIC}_{C_K} = NS_K + V_K + \frac{V_K(V_K+1)}{N-V_K-1}$ and $V_K = \dim(\lambda_K) - K + \dim(\gamma_K) - n = (n-1)K + n(K-1)$.

- Step 3: Compute the average latent entropy and the average latent dimension as the following expectation values:

$$\bar{S} = \sum_{K=1}^n p_K S_K, \quad \text{and} \quad \bar{K} = \sum_{K=1}^n p_K K. \quad (\text{Equation 4})$$

The latent dimension and latent entropy are, thus, calculated by assuming that the transition $X(t) \rightarrow X(t+1)$ may happen through any number of latent dimensions, $K = 1, \dots, n$, and then weighting them with their posterior probability to occur. This can be seen in analogy to the concept of path integrals where the transition amplitude of a particle from an initial to a final state is computed as a weighted sum of all possible trajectories (Feynman, 1948; Feynman et al., 2010; Kleinert, 2009).

The average latent dimension K quantifies how many latent states best describe each pixel’s underlying dynamics. The smallest number of latent variables is 1, which corresponds to no memory at all, i.e., it does not matter which is the initial state since they are all taken to a single latent intermediate state. A higher K indicates a higher dependence on the initial states. The average latent entropy S quantifies the stochasticity of each pixel’s underlying dynamics. A low S means that the system is very predictable. A higher S is associated to the increase in the randomness of the underlying dynamics.

The latent measures are calculated for each pixel individually; therefore, the iteration step in the calculation depends neither on the pixel number nor on the statistical size N as long as $N > n^2$. The computation and memory cost depend only on the maximal number of latent dimensions n and scale as $\mathcal{O}(n^4)$ and $\mathcal{O}(n^2)$, respectively.

Linear Markovian “falling toast” model:

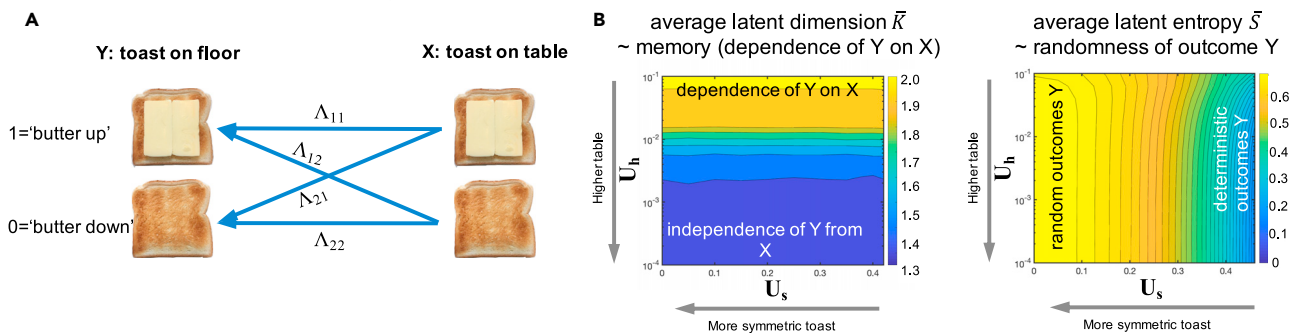


Figure 2. An explanatory toy example of the falling toast experiment

(A) Visualization of transition probabilities between initial and final state;

(B) Results for the latent dimension and latent entropy for different values of the two parameters of the model corresponding to the toast's asymmetry U_s and table height U_h .

For comparison, the computational costs of the GMM methods mentioned in the [Gaussian mixture model](#) subsection, for a given observational data of dimension D with N time frames and n possible Gaussian distributions (i.e. latent states), are to the leading order as $\mathcal{O}(n^2 ND)$ and require $\mathcal{O}(n(N + D))$ of memory (Horenko et al., 2019).

We would like to note that this method works not only for considering each pixel individually but also when joining pixels into pixel patches. While bigger patches allow a deeper analysis of latent processes and long-range interactions, they cause loss of accuracy for pattern recognition. For example, joining all pixels to one block would remove the possibility to recognize any patterns. For each case, considering the data acquisition's sensitivity and resolution, there is an optimal pixel patch size. Nevertheless, analyzing the pixels independently is accurate for strong local interactions, particularly when each pixel can assume various values.

In the following, we will present a minimal model that explains the physical ideas of the latent entropy and latent memory.

Minimal model—a linear Markovian falling toast model

As a toy example, let us consider the study of “Murphy's law”, for which Robert Matthews obtained the IG Nobel Prize in 1996 (Matthews, 1995). In particular, Matthews demonstrated that it is not by implementation of the mysterious Murphy's law that a toast falls mostly on the buttered side, but it is instead a consequence of the latent effect of the standard table height from which the toast is falling. This is reflected in what we refer in this article as the average latent dimension K (Equation 4). It encodes information about the degree to which the outcome depends on the initial configuration. As the typical table height influences the outcome, this means that the latent dimension is larger than one, i.e., the final state has a strong memory of the initial state. Generally, however, a fully symmetric toast falling from a sufficiently high table has on average the same chance of falling on either of the sides when sampling various starting configurations. This represents a maximal latent entropy S (Equation 4) configuration. The latter can be reduced by favoring one of the sides, e.g., by spreading butter on one side. To make the notation used in the previous section even more explicit, we will describe the falling toast experiment in terms of its corresponding linear Markovian model.

In this experiment, the initial and final states, $X(t = 0) \equiv X$ and $X(t = 1) \equiv Y$, respectively, correspond to the orientation of the toast before and after falling from the table. The two possible configurations for X and Y are “butter up” which we will label by 1 and “butter down” which we will label by 0, see Figure 2A. The transition probability is given as follows:

$$\begin{bmatrix} \mathbb{P}[Y = 1] \\ \mathbb{P}[Y = 0] \end{bmatrix} \equiv \Lambda \begin{bmatrix} \mathbb{P}[X = 1] \\ \mathbb{P}[X = 0] \end{bmatrix}, \quad (\text{Equation 5})$$

where Λ is the transition matrix given as follows:

$$\Lambda = \begin{bmatrix} \mathbb{P}[Y=1|X=1] & \mathbb{P}[Y=1|X=0] \\ \mathbb{P}[Y=0|X=1] & \mathbb{P}[Y=0|X=0] \end{bmatrix}. \quad (\text{Equation 6})$$

According to the law of total probability, the columns of the transition matrix sum to 1. This implies that there are only 2° of freedom for this 2 × 2 matrix. We assign to these degrees of freedom the quantities U_s and U_h as deviations from the transition matrix corresponding to the Bernoulli model, where all probabilities are equal to 0.5, i.e.,

$$\Lambda = \begin{bmatrix} 0.5 - U_s + U_h & 0.5 - U_s - U_h \\ 0.5 + U_s - U_h & 0.5 + U_s + U_h \end{bmatrix}. \quad (\text{Equation 7})$$

The parameter U_h (h for height) controls the randomizing effect due to the height of the table—as well as the amount of memory of the variable Y (toast on the floor) and its dependence on the variable X (toast on the table). The parameter U_s (s for symmetry) models a change of the asymmetry of the toast. Note that only combinations of parameters U_h and U_s are allowed such that all matrix entries are between zero and 1.

We consider three limiting cases to better understand the roles of U_h and U_s .

1. For a symmetric toast falling from a high table, we have that $U_h = U_s = 0$; therefore, the following equation is obtained:

$$\Lambda = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}. \quad (\text{Equation 8})$$

In this case, it is impossible to predict which face of the toast faces up when at the floor, i.e., it is independent of the initial state and both outcomes are equally probable. Thus, the system has no memory ($K = 1$) and no predictability (S is maximum). This corresponds to the Bernoulli experiment.

2. For a symmetric toast falling from a very low table (such that the toast has no chance to flip), we have that $U_s = 0$ and $U_h = 0.5$; therefore, the following equation is obtained:

$$\Lambda = \begin{bmatrix} 1.0 & 0.0 \\ 0.0 & 1.0 \end{bmatrix}. \quad (\text{Equation 9})$$

In this system, given an initial state, we know for sure the final state, i.e., the memory is maximal (K is maximum, here $K = 2$) and low predictability (S is maximum).

3. For a very asymmetric toast falling from a high table, we have that $U_s = 0.5$ and $U_h = 0$; therefore, the following equation is obtained:

$$\Lambda = \begin{bmatrix} 0.0 & 0.0 \\ 1.0 & 1.0 \end{bmatrix}. \quad (\text{Equation 10})$$

Asymmetric toast means that the center of mass is shifted in such a way that the toast tends to have the buttered face down while the high table means that it has time to flip, if necessary. The final state is, therefore, independent of the initial state. This means that the system has no memory ($K = 1$), but high predictability ($S = 0$), as the outcome is clear.

After 100 random experiments according to the model in Equation 7, for every combination of the parameters U_s and U_h , we calculated the expected values of the average latent entropy and average latent dimension, see Figure 2B. We find that the average latent dimension depends only on the parameter U_h while it is independent of U_s , whereas for the latent entropy, it is the other way around. This can be understood in the following way: the parameter U_h determines the correlation between an initial and a final state, i.e., the memory of the system. The smaller U_h , the more deterministic is the system's behavior, meaning the fewer latent processes appear and thus the smaller the latent dimension. Increasing U_s , on the contrary, increases

Microscopy of mouse brain

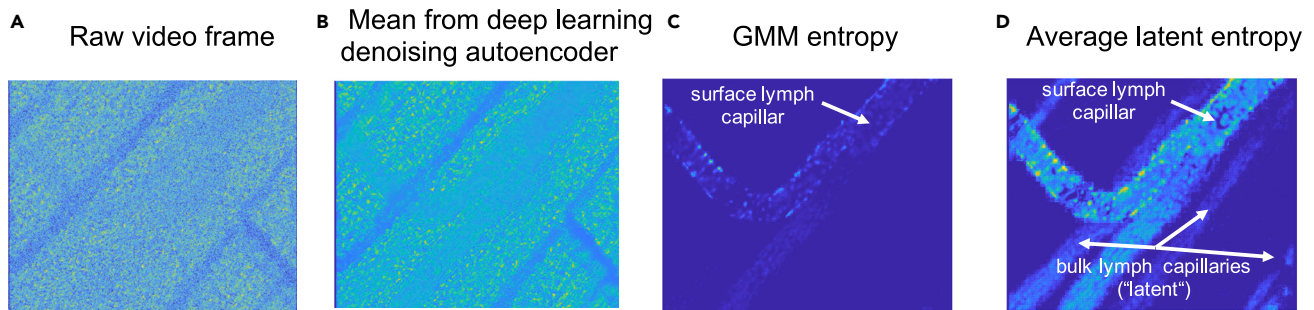


Figure 3. Analysis of the microscopy video of lymph flow in a mouse brain

(A–D) (A) Raw video frame, (B) results obtained by calculating the mean from the deep learning denoising autoencoder, (C) analysis by means of the GMM entropy clearly revealing surface lymph capillaries, (D) analysis by means of the latent entropy revealing even bulk lymph capillaries.

the predictability of the outcome and therefore leads to a smaller latent entropy, while not influencing the system's memory. Figure 2B indicates that the latent dimension and latent entropy are two rather orthogonal measures revealing very different types of information content.

APPLICATIONS

In this section, we will exemplify the strength of latent measures, focusing on the latent entropy, by presenting the analysis of data from three different fields in natural science belonging to different space and time scales. First, we apply the methodology to biological systems, specifically the microscopy video of a lymph flow in a mouse brain. We demonstrate that the latent entropy measure reveals previously poorly observed parts of the glymphatic network system (Iliff et al., 2012; Begley, 2012). We will then present an application to analysis of video data from micromagnetic experiments, revealing the non-trivial latent features in the underlying sample. We show that the high accuracy in examining the underlying dynamics allows for detecting latent entropy gradients in the magnetic samples. In the analyzed data, we associate the detected gradient patterns to a variation of temperature along the sample. As a third example, we apply the latent tools to amateur astronomical observations to demonstrate that even with low sensitivity of the particular astronomic instruments and significant atmosphere-induced noise, it is possible to detect subtle features that are otherwise only directly accessible to much more powerful instruments.

Glymphatic system of a mouse brain

In this first example, we show the power of the latent measure to observe the glymphatic system of a mouse brain. The glymphatic system in the brain was unobserved for a long time because the fluid in the glymphatic network is transparent. Until 2012, there was no solid evidence for its existence (Iliff et al., 2012). With the use of advanced latent measures, however, one can visually observe the flux corresponding to the glymphatic network from microscopic video of a mouse brain, see Figure 3.

The glymphatic system constitutes an autonomous lymph transport network in the brain, responsible for the disposal of waste products during the sleeping phase (Iliff et al., 2012; Begley, 2012; Nedergaard, 2013; Benveniste et al., 2019). Deeply understanding its functionality is particularly important in the study of neurodegenerative diseases, which are associated with an atypical accumulation of cellular waste products (Benveniste et al., 2019).

Figure 3 summarizes the analysis results for the light microscopy video of the living mouse brain tissue with a flow of the transparent lymph fluid through the capillaries of the glymphatic system. The capillaries of the glymphatic system are not directly visible either in the raw video data, see Figure 3A, or in the mean obtained by using a commercial deep learning denoising autoencoder from the “Deep Learning Toolbox” of MATLAB (Zhang et al., 2017; Berghout et al., 2020, 2021), see Figure 3B. Application of the common GMM entropy (Zoran and Weiss, 2011; Bouman et al., 2018; The Event Horizon Telescope Collaboration, 2019a; Greggio et al., 2012) does allow to visualize only surface capillaries, see Figure 3C. The deeper lying bulk capillaries are not directly visible and can only be revealed through latent effects. As can be seen from

Thermal movement of skyrmions without external field

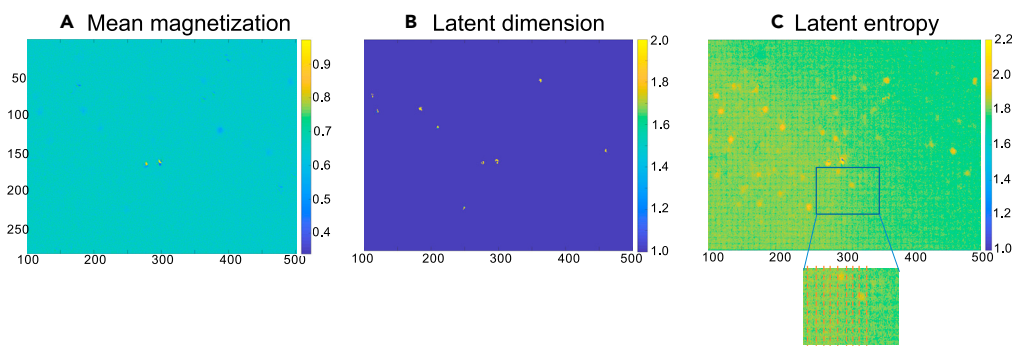


Figure 4. Data analysis of a MOKE experiment

The axes denote the number of pixels and the color code represents (A) the mean value of the out-of-plane magnetization, (B) the latent dimension, and (C) the latent entropy. The inset elucidates the superstructure grid pattern imposed by the video compression.

the Figure 3D, applying the latent entropy measure one can extract and visualize the capillary pattern, including the bulk ones.

Temperature gradient in magnetization experiment

In this second example, we show that the high sensitivity of the latent measures allows detecting subtle material inhomogeneities by analyzing the video data from micromagnetic measurements. We apply the latent tools to a video of the magnetization configuration in specially tailored low-pinning, multilayer material Ta(5nm)/Co₂₀Fe₂₀B₂₀(1nm)/Ta(0.08nm)/MgO(2nm)/Ta(5nm) stacks (Zázvorka et al., 2019) obtained by MOKE microscopy (Hubert & Schäfer, 2009; Huang et al., 1994). In this experiment, for each resolution pixel, only the out-of-plane component of the magnetization is detected at about room temperature (Note that 305 pixels correspond to 50 μm). The time step between two measurements, i.e., two video frames, is 62.5 ms. The experiment in Figure 4 aimed at studying rather homogeneous materials, striving for a free motion of magnetic skyrmions and avoiding impurities where magnetic textures get pinned. The time record of the skyrmions' positions, however, revealed that there are preferred positions where the skyrmions tend to stay longer and which can indirectly be associated to the existence of inhomogeneities (Zázvorka et al., 2019). These inhomogeneities strongly influence, for example, the temperature dependence of the skyrmion diffusion coefficient (Zázvorka et al., 2019), and thus, it is important to detect even weak or small material defects. Strong material inhomogeneities can be resolved by simple means such as the mean value of the magnetization, as shown in Figure 4A. The average latent dimension (see Figure 4B) sharply identifies impurities where skyrmions are more likely to be pinned, e.g., where the skyrmions have a longer memory of their previous state. The latent entropy (see Figure 4C) not only reveals material inhomogeneities as spots of higher entropy but also allows us to discern two other significant features: an entropy gradient across the sample and a subtle grid pattern superstructure, as shown in the inset. The latter is induced by the video compression applied to the experimental data, as discussed in the study by (Horenko et al., 2019).

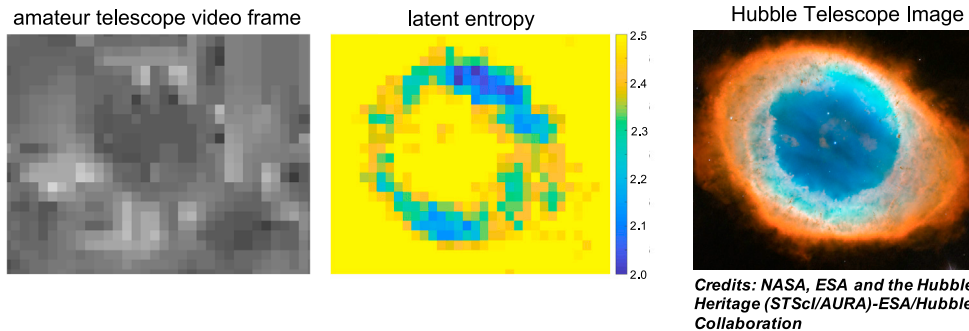
The gradient in the latent entropy is associated with a field gradient in the underlying magnetization dynamics, and the experimental conditions are consistent with the interpretation that it corresponds to a temperature gradient in the sample. This opens up a path toward analyzing the temperature distribution across devices, which are typically very hard to detect by means of common magnetic imaging techniques.

Astronomical observations

In this third example, we show how the latent tools can be used to extract latent features from noisy amateur astronomy videos. We apply the latent tools explained above to remove the noise induced by atmospheric fluctuations and to detect the underlying features by identifying the dynamical patterns in the video data.

Analysis of the amateur videos of astronomical observations

A Telescope video of the ring nebula in Lira (M57, several lightyears scale)



B Telescope video of the star cluster in Hercules (M11, 10'000 lightyears scale)

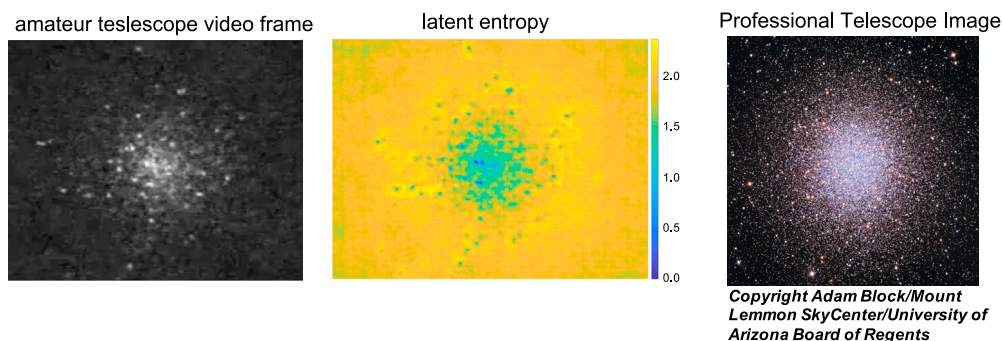


Figure 5. Analysis of the amateur videos of astronomical observations (A) for the M57 Ring Nebula in Lira and (B) for the M31 star cluster in Hercules.

Figure 5 shows results obtained for two amateur infrared videos: (a) Ring Nebula in Lira (M57) and (b) a star cluster in Hercules (M11), comparing the patterns obtained with the latent measures to the direct observations with the Hubble Space Telescope. We notice that despite of the apparent strong atmospheric fluctuations, these two videos can still be used to extract the underlying image patterns of a remarkable quality. Comparing the frames from the raw video data in the first column of Figure 5 with the latent entropy measure (middle column), one can see that the number of details, for instance, the star concentration, is significantly increased such that it becomes much closer to the features obtained with the Hubble telescope, as shown in the third column of Figure 5.

Computational and statistical methods to identify subtle features in the noisy video data are of large importance for astronomical observations and allowed for important discoveries in the past, such as the cosmic microwave background radiation, gravitational waves, existence and motion of exoplanets, and imaging of black holes (Zoran and Weiss, 2011; Greggio et al., 2012; Bouman et al., 2018; The Event Horizon Telescope Collaboration, 2019b). The high accuracy and the low computational cost of the latent measures may potentially help to enhance a detection of weak latent features in these application domains.

DISCUSSION AND CONCLUSION

We have shown the utility of advanced data analysis tools in application to problems of pattern extraction and denoising across a wide range of spatiotemporal scales, from nanoscales to astronomic length scales, discussing also potential consequences for disparate research fields ranging across physics and biology. Particular emphasis was placed on the recently introduced data analysis tools—on the latent entropy and the latent dimension—aiming at disentangling the latent effects induced by predictability and memory in the observed data dynamics. Operating beyond common assumptions like Gaussianity and homogeneity and taking into account the temporal ordering of the data, it was shown that these tools can reveal subtle latent features that are otherwise invisible to popular latent inference methods like GMMs and deep learning denoising autoencoders. In particular, these tools also avoid overfitting issues as they build an

expectation over all possible non-parametric (i.e., without explicit parametric assumptions like Gaussianity) latent fits with latent dimension K spanning the whole range from $K = 1$ to $K=N$. The calculation of expectation values is based on the posterior Akaike weights—one of the popular tools of information theory that were designed to avoid overfitting.

A common caveat for all of the data-driven measures compared in this manuscript is that even though these methods can help to identify different patterns hidden in the data, they do not directly provide means to determine the precise sources of these different patterns. They do, however, reveal interesting and even subtle features and inhomogeneities that can then be further investigated.

To summarize, machine learning and data analysis tools will play an increasingly important role in various application areas—and exploiting advanced tools that aim at detecting latent dynamical patterns in time-resolved data promises to achieve much deeper insights into the nature of the underlying phenomena.

ACKNOWLEDGMENTS

We would like to thank the group of M. Kläui, in particular J. Zazvorká, for discussions and contributing with the experimental data used in Figure 4. We are also grateful for the input from J. Sinova. We acknowledge funding from the Emergent AI Center funded by the Carl-Zeiss-Stiftung. K.E.S. acknowledges funding from the German Research Foundation (DFG) project No. 320163632 (EV 196/2-1). The work of I.H. was funded by the Mercator Fellowship in the DFG Collaborative Research Center 1114 “Scaling Cascades in Complex Systems”.

AUTHOR CONTRIBUTIONS

D.R.R., K.E.-S., and I.H. wrote the paper. I.H. performed the analysis on the data. All authors contributed to designing the study, interpreted the data, read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Abbott, B.P., Abbott, R., Abbott, T.D., Acernese, F., et al. (2017). Gw170817: observation of gravitational waves from a binary neutron star inspiral. *Phys. Rev. Lett.* 119, 161101, <https://doi.org/10.1103/PhysRevLett.119.161101>.
- Al-Jarrah, O.Y., Yoo, P.D., Muhaidat, S., Karagiannidis, G.K., and Taha, K. (2015). Efficient machine learning for big data: a review. *Big Data Res.* 87–93, <https://doi.org/10.1016/j.bdr.2015.04.001>.
- Barndorff-Nielsen, O.E., and Shephard, N. (2001). Non-Gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc. Ser. B (Statistical Methodology)* 63, 167–241, <https://doi.org/10.1111/1467-9868.00282>.
- Bauer, D.J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behav. Res.* 42, 757–786, <https://doi.org/10.1080/00273170701710338>.
- Bayes, T. (1763). LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. (*Philosophical Transactions of the Royal Society of London*), pp. 370–418, <https://doi.org/10.1098/rstl.1763.0053>.
- Begley, D.J. (2012). Brain superhighways. *Sci. Translational Med.* 4, 147, <https://doi.org/10.1126/scitranslmed.3004611>.
- Bellone, E., Hughes, J., and Guttorp, P. (2000). A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Clim. Res.* 1–12, <https://doi.org/10.3354/cr015001>.
- Bengio, Y., and LeCun, Y. (2007). Scaling learning algorithms towards AI. In *Large-Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste, and J. Weston, eds. (MIT Press). <https://nyuscholars.nyu.edu/en/publications/scaling-learning-algorithms-towards-ai>.
- Benveniste, H., Liu, X., Koundal, S., Sanggaard, S., Lee, H., and Wardlaw, J. (2019). The glymphatic system and waste clearance with brain aging: a review. *Gerontology*, 106–119, <https://doi.org/10.1159/000490349>.
- T. Berghout, Denoising Autoencoder (2021) (<https://www.mathworks.com/matlabcentral/fileexchange/71115-denoising-autoencoder>), MATLAB Central File Exchange.
- Berghout, T., Mouss, L.-H., Kadri, O., Saïdi, L., and Benbouzid, M. (2020). Aircraft Engines Remaining Useful Life Prediction with an Adaptive Denoising Online Sequential Extreme Learning Machine, *Engineering Applications of Artificial Intelligence* 103936, <https://doi.org/10.1016/j.engappai.2020.103936>.
- Blum, A.L., and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artif. Intelligence*, 245–271, [https://doi.org/10.1016/S0004-3702\(97\)00063-5](https://doi.org/10.1016/S0004-3702(97)00063-5).
- Bouman, K.L., Johnson, M.D., Dalca, A.V., Chael, A.A., Roelofs, F., Doeleman, S.S., and Freeman, W.T. (2018). Reconstructing video of time-varying sources from radio interferometric measurements. *IEEE Trans. Comput. Imaging* 4, 512–527, <https://doi.org/10.1109/TCI.2018.2838452>.
- Brailean, J., Kleihorst, R., Efstratiadis, S., Katsaggelos, A., and Lagendijk, R. (1995). Noise reduction filters for dynamic image sequences: a review. *Proc. IEEE*, 1272–1292, <https://doi.org/10.1109/5.406412>.
- Buades, A., Coll, B., and Morel, J.-M. (2005). A non-local algorithm for image denoising. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) (IEEE), pp. 60–65, <https://doi.org/10.1109/CVPR.2005.38>.
- Collier, N., Nobata, C., and Tsujii, J.-i. (2000). Extracting the names of genes and gene products with a hidden Markov model. In *Proceedings of the 18th Conference on Computational Linguistics -, COLING '00* (Association for Computational Linguistics), pp. 201–207, <https://doi.org/10.3115/990820.990850>.

- Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* 2080–2095, <https://doi.org/10.1109/TIP.2007.901238>.
- Deng, C., Ji, X., Rainey, C., Zhang, J., and Lu, W. (2020). Integrating machine learning with human knowledge. *iScience*, 101656, <https://doi.org/10.1016/j.isci.2020.101656>.
- Dias, J.G., Vermunt, J.K., and Ramos, S. (2015). Clustering financial time series: new insights from an extended hidden Markov model. *Eur. J. Oper. Res.* 3, 852–864, <https://doi.org/10.1016/j.ejor.2014.12.041>.
- Ding, C., Li, T., and Peng, W. (2006). Nonnegative matrix factorization and probabilistic latent semantic indexing: equivalence, chi-square statistic, and a hybrid method. In *Proceedings of the 21st National Conference on Artificial Intelligence, 1Proceedings of the 21st National Conference on Artificial Intelligence (AAAI'06)*, pp. 342–347.
- Enßlin, T.A., Frommert, M., and Kitaura, F.S. (2009). Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Phys. Rev. D* 10, 105005, <https://doi.org/10.1103/PhysRevD.80.105005>.
- Ernst, R.R. (1966). *Sensitivity Enhancement in Magnetic Resonance* (Academic Press), pp. 1–135, <https://doi.org/10.1016/B978-1-4832-3115-0.50008-9>.
- Feynman, R.P. (1948). Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* 20, 367–387, <https://doi.org/10.1103/RevModPhys.20.367>.
- Feynman, R.P., Hibbs, A.R., and Styer, D.F. (2010). *Quantum Mechanics and Path Integrals* (Courier Corporation) 978-0486477220.
- François, O., and Durand, E. (2010). Spatially explicit bayesian clustering models in population genetics. *Mol. Ecol. Resour.* 10, 773–784, <https://doi.org/10.1111/j.1755-0998.2010.02868.x>.
- Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models* (Springer) 978-0387357683.
- Gardiner, H. (2004). *Handbook of Stochastic Methods* (Springer) 978-3540707127.
- Gerber, S., and Horenko, I. (2015). Improving clustering by imposing network information. *Sci. Adv.* 1, e1500163, <https://doi.org/10.1126/sciadv.1500163>.
- Gerber, S., and Horenko, I. (2017). Toward a direct and scalable identification of reduced models for categorical processes 114, 4863–4868, <https://doi.org/10.1073/pnas.1612619114>.
- Gerber, S., Olson, S., Noe, F., and Horenko, I. (2018). A scalable approach to the computation of invariant measures for high-dimensional Markovian systems. *Scientific Rep.* 8, 1796, <https://doi.org/10.1038/s41598-018-19863-4>.
- Gerber, S., Pospisil, L., Navandar, M., and Horenko, I. (2020). Low-cost scalable discretization, prediction, and feature selection for complex systems. *Sci. Adv.* 6, eaaw0961, <https://doi.org/10.1126/sciadv.aaw0961>.
- Goyal, B., Dogra, A., Agrawal, S., Sohi, B., and Sharma, A. (2020). Image denoising review: from classical to state-of-the-art approaches. *Inf. Fusion*, 220–244, <https://doi.org/10.1016/j.inffus.2019.09.003>.
- Greggio, N., Bernardino, A., Laschi, C., Dario, P., and Santos-Victor, J. (2012). Fast estimation of Gaussian mixture models for image segmentation. *Machine Vis. Appl.* 23, 773–789, <https://doi.org/10.1007/s00138-011-0320-5>.
- Hassan, M., and Nath, B. (2005). Stock market forecasting using hidden Markov model: a new approach. In *5th International Conference on Intelligent Systems Design and Applications (ISDA'05)* (IEEE), pp. 192–196, <https://doi.org/10.1109/ISDA.2005.85>.
- Hinton, G.E. (2006). Reducing the dimensionality of data with neural networks. *Science*, 504–507, <https://doi.org/10.1126/science.1127647>.
- Hipp, J.R., and Bauer, D.J. (2006). Local solutions in the estimation of growth mixture models. *Psychol. Methods* 11 (1), 36, <https://doi.org/10.1037/1082-989X.11.1.36>.
- Hofmann, T. (1999). Probabilistic latent semantic indexing. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, '99 (SIGIR), pp. 50–57.
- Hofmann, T. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learn.* 42, 177–196, <https://doi.org/10.1023/A:1007617005950>.
- Horenko, I. (2020). On a scalable entropic breaching of the overfitting barrier for small data problems in machine learning. *Neural Comput.* 1563–1579, https://doi.org/10.1162/neco_a_01296.
- Horenko, I., Rodrigues, D., O'Kane, T., and Everschor-Sitte, K. (2019). Scalable detection of latent patterns across scales – from nanomagnets to astrophysics. *arXiv* 1907, 04601.
- Huang, F., Kief, M.T., Mankey, G.J., and Willis, R.F. (1994). Magnetism in the few-monolayers limit: a surface magneto-optic kerr-effect study of the magnetic behavior of ultrathin films of co, ni, and co-ni alloys on cu(100) and cu(111). *Phys. Rev. B* 49, 3962–3971, <https://doi.org/10.1103/PhysRevB.49.3962>.
- Hubert, A., and Schäfer, R. (2009). *Magnetic Domains*, 3rd Edition (Springer Berlin Heidelberg). <https://doi.org/10.1007/978-3-540-85054-0>.
- Hurvich, C.M., and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* 76, 297–307, <https://doi.org/10.1093/biomet/76.2.297>.
- Iliff, J.J., Wang, M., Liao, Y., Plogg, B.A., Peng, W., Gundersen, G.A., Benveniste, H., Vates, G.E., Deane, R., Goldman, S.A., et al. (2012). A paravascular pathway facilitates csf flow through the brain parenchyma and the clearance of interstitial solutes, including amyloid β . *Sci. Transl. Med.* 4, 147, <https://doi.org/10.1126/scitranslmed.3003748>.
- Jurafsky, D., and Martin, J.H. (2009). *Speech and Language Processing*, 2nd Edition (Prentice-Hall, Inc.).
- KaewTraKulPong, P., and Bowden, R. (2002). An improved adaptive background mixture model for real-time tracking with shadow detection. In *Video-based Surveillance Systems* (Springer), pp. 135–144, https://doi.org/10.1007/978-1-4615-0913-4_11.
- Kleinert, H. (2009). *Path Integrals in Quantum Mechanics, Statistics, Polymer Physics, and Financial Markets* (World scientific). <https://doi.org/10.1142/7305>.
- Loh, P.-R., Tucker, G., Bulik-Sullivan, B.K., Vilhjálmsson, B.J., Finucane, H.K., Salem, R.M., Chasman, D.I., Ridker, P.M., Neale, B.M., Berger, B., et al. (2015). Efficient bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* 47, 284, <https://doi.org/10.1038/ng.3190>.
- Luisier, F., Blu, T., and Unser, M. (2010). SURE-LET for Orthonormal Wavelet-Domain Video Denoising (IEEE Transactions on Circuits and Systems for Video Technology), pp. 913–919, <https://doi.org/10.1109/TCSVT.2010.2045819>.
- Luisier, F., Blu, T., and Unser, M. (2011). Image Denoising in Mixed Poisson-Gaussian Noise (IEEE Transactions on Image Processing), pp. 696–708, <https://doi.org/10.1109/TIP.2010.2073477>.
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., and Zisserman, A. (2009). Non-local sparse models for image restoration. In *2009 IEEE 12th International Conference on Computer Vision (IEEE)*, pp. 2272–2279, <https://doi.org/10.1109/ICCV.2009.5459452>.
- Manogaran, G., Vijayakumar, V., Varatharajan, R., Malarvizhi Kumar, P., Sundarasekar, R., and Hsu, C.-H. (2018). Machine learning based big data processing framework for cancer diagnosis using hidden Markov model and GM clustering. *Wireless Personal. Commun.* 2099–2116, <https://doi.org/10.1007/s11277-017-5044-z>.
- Matthews, R.A.J. (1995). Tumbling toast, Murphy's Law and the fundamental constants. *Eur. J. Phys.* 16, 172, <https://doi.org/10.1088/0143-0807/16/4/005>.
- McKenna, S.J., Raja, Y., and Gong, S. (1998). Object tracking using adaptive colour mixture models. In *Asian Conference on Computer Vision* (Springer), pp. 615–622, https://doi.org/10.1007/3-540-63930-6_174.
- Melnykov, V., and Maitra, R. (2010). Finite mixture models and model-based clustering. *Stat. Surv.* 4, 80–116, <https://doi.org/10.1214/09-SS053>.
- Narasimhan, V., Danecek, P., Scally, A., Xue, Y., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/ RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. *Bioinformatics* 11, 1749–1751, <https://doi.org/10.1093/bioinformatics/btw044>.
- Nedergaard, M. (2013). Garbage truck of the brain. *Science* 6140, 1529–1530, <https://doi.org/10.1126/science.1240514>.
- Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z.B., and Swami, A. (2017). Practical

black-box Attacks against machine learning. In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ACM), pp. 506–519, <https://doi.org/10.1145/3052973.3053009>.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. *J. Machine Learn. Res.* 12, 2825–2830.

Penzias, A.A., and Wilson, R.W. (1965). A measurement of excess antenna temperature at 4080 mc/s. *Astrophysical J.* 142, 419–421, <https://doi.org/10.1086/148307>.

Phillips, P.C., and Loretan, M. (1991). Estimating long-run economic equilibria. *Rev. Econ. Stud.* 58, 407–436, <https://doi.org/10.2307/2298004>.

Pinto, R.C., and Engel, P.M. (2015). A fast incremental Gaussian mixture model. *PLoS One* 10, e0139931, <https://doi.org/10.1371/journal.pone.0139931>.

K.K. Rachuri, M. Musolesi, C. Mascolo, P.J. Rentfrow, C. Longworth, A. Aucinas, 2010, Emotionsense: a mobile phones based adaptive platform for experimental social psychology research, in: Proceedings of the 12th ACM international conference on Ubiquitous computing. 281–290. doi:10.1145/1864349.1864393.

Reynolds, D.A. (1995). Speaker identification and verification using Gaussian mixture speaker models. *Speech Commun.* 17, 91–108.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Machine Intelligence*, 206–215, <https://doi.org/10.1038/s42256-019-0048-x>.

Runge, J., Petoukhov, V., Donges, J.F., Hlinka, J., Jajcay, N., Vejmelka, M., Hartman, D., Marwan, N., Paluš, M., and Kurths, J. (2015). Identifying causal gateways and mediators in complex spatio-temporal systems. *Nat. Commun.* 1, 8502, <https://doi.org/10.1038/ncomms9502>.

Rupe, A., Prabhat, M., Crutchfield, J.P., Kumar, N., Epifanov, V., Kashinath, K., Pavlyk, O., Schlimbach, F., Patwary, M., Maidanov, S., and Lee, V. (2019). DisCo: physics-based unsupervised discovery of coherent structures in spatiotemporal systems. In IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC), pp. 75–87, <https://doi.org/10.1109/MLHPC49564.2019.00013>.

Santosh, D.H.H., Venkatesh, P., Poornesh, P., Rao, L.N., and Kumar, N.A. (2013). Tracking multiple moving objects using Gaussian mixture model. *Int. J. Soft Comput. Eng. (Ijsce)* 3, 114–119.

Schuller, B., Rigoll, G., and Lang, M. (2003). Hidden Markov model-based speech emotion recognition. In 2003 IEEE International Conference on Acoustics, Speech, and Signal

Processing. Proceedings. (ICASSP 03) (IEEE). II–1–4. <https://doi.org/10.1109/ICASSP.2003.1202279>.

Shannon, C.E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.* 27, 379–423, <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

Shao, L., Yan, R., Li, X., and Liu, Y. (2013). From heuristic optimization to dictionary learning: a review and comprehensive comparison of image denoising algorithms. *IEEE Trans. Cybernetics*, 1001–1013, <https://doi.org/10.1109/TCYB.2013.2278548>.

Shao, Ling, Zhang, Hui, and de Haan, G. (2008). An overview and performance evaluation of classification-based least squares trained filters. *IEEE Trans. Image Process.* 10, 1772–1782, <https://doi.org/10.1109/TIP.2008.2002162>.

Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* 19 (Suppl 2), ii215–ii225, <https://doi.org/10.1093/bioinformatics/btg1080>.

Stauffer, C., and Grimson, W.E.L. (1999). Adaptive background mixture models for real-time tracking. In Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. PR00149), Vol. 2 Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No. PR00149) (IEEE), pp. 246–252, <https://doi.org/10.1109/CVPR.1999.784637>.

Stockman, G., and Shapiro, L.G. (2001). *Computer Vision*, 1st Edition (Prentice Hall PTR) 978-0130307965.

Tanner, M.A. (2012). *Tools for Statistical Inference* (Springer) 978-1461240242.

The Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczkó, A.K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al. (2019a). First M87 event horizon telescope results. I. The shadow of the supermassive black hole. *Astrophysical J.* 875, L4, <https://doi.org/10.3847/2041-8213/ab0ec7>.

The Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., Alef, W., Asada, K., Azulay, R., Baczkó, A.K., Ball, D., Baloković, M., Barrett, J., Bintley, D., et al. (2019b). First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *Astrophysical J.* 875, L4, <https://doi.org/10.3847/2041-8213/ab0e85>.

P.A. Torres-Carrasquillo, E. Singer, M.A. Kohler, R.J. Greene, D.A. Reynolds, J.R. Deller Jr, 2002, Approaches to language identification using gaussian mixture models and shifted delta cepstral features, in: Seventh international conference on spoken language processing, .

Turelli, M., and Barton, N.H. (1994). Genetic and statistical analyses of strong selection on

polygenic traits: what, me normal? *Genetics* 138, 913–941, PMID:PMC1206238.

Vesselinov, V., Mudunuru, M., Karra, S., O'Malley, D., and Alexandrov, B. (2019). Unsupervised machine learning based on non-negative tensor factorization for analyzing reactive-mixing. *J. Comput. Phys.* 85–104, <https://doi.org/10.1016/j.jcp.2019.05.039>.

Wiener, N. (1950). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications* (MIT press) 978-0262730051.

Williams, M.O., Kevrekidis, I.G., and Rowley, C.W. (2015). A data-driven approximation of the koopman operator: extending dynamic mode decomposition. *J. Nonlinear Sci.* 1307–1346, <https://doi.org/10.1007/s00332-015-9258-5>.

Wink, A.M., and Roerdink, J.B.T.M. (2004). Denoising Functional MR Images: A Comparison of Wavelet Denoising and Gaussian Smoothing23 (IEEE Transactions on Medical Imaging), pp. 374–387, <https://doi.org/10.1109/TMI.2004.824234>.

Xi, L., Fondufe-Mittendorf, Y., Xia, L., Flatow, J., Widom, J., and Wang, J.-P. (2010). Predicting nucleosome positioning using a duration Hidden Markov Model. *BMC Bioinformatics*, 346, <https://doi.org/10.1186/1471-2105-11-346>.

Ye, H., Beamish, R.J., Glaser, S.M., Grant, S.C.H., Hsieh, C.-h., Richards, L.J., Schnute, J.T., and Sugihara, G. (2015). Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. *Proc. Natl. Acad. Sci.* 113, E1569–E1576, <https://doi.org/10.1073/pnas.1417063112>.

Zázvorka, J., Jakobs, F., Heinze, D., Keil, N., Kromin, S., Jaiswal, S., Litzius, K., Jakob, G., Virnau, P., Pinna, D., et al. (2019). Thermal skyrmion diffusion used in a reshuffler device. *Nat. Nanotechnol.* 14, 658–661, <https://doi.org/10.1038/s41565-019-0436-8>.

Zenil, H., Kiani, N.A., Zea, A.A., and Tegnér, J. (2019). Causal deconvolution by algorithmic generative models. *Nat. Machine Intelligence*, 58–66, <https://doi.org/10.1038/s42256-018-0005-0>.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian Denoiser: Residual Learning of Deep CNN for Image Denoising (IEEE Transactions on Image Processing), pp. 3142–3155, <https://doi.org/10.1109/TIP.2017.2662206>.

Zhang, K., Zuo, W., Chen, Y., Meng, D., and Zhang, L. (2017). Beyond a Gaussian denoiser: residual learning of deep cnn for image denoising. *Trans. Img. Proc.* 26, 3142–3155, <https://doi.org/10.1109/TIP.2017.2662206>.

Zoran, D., and Weiss, Y. (2011). From learning models of natural image patches to whole image restoration. In 2011 International Conference on Computer Vision, pp. 479–486, <https://doi.org/10.1109/ICCV.2011.6126278>.