

Constructing Compact Causal Mathematical Models for Complex Dynamics

Yuankun Xue

University of Southern California
Los Angeles, CA 90089
yuankunx@usc.edu

Paul Bogdan

University of Southern California
Los Angeles, CA 90089
pbogdan@usc.edu

ABSTRACT

From microbial communities, human physiology to social and biological/neural networks, complex interdependent systems display multi-scale spatio-temporal patterns that are frequently classified as non-linear, non-Gaussian, non-ergodic, and/or fractal. Distinguishing between the sources of nonlinearity, identifying the nature of fractality (space versus time) and encapsulating the non-Gaussian characteristics into dynamic causal models remains a major challenge for studying complex systems. In this paper, we propose a new mathematical strategy for constructing compact yet accurate models of complex systems dynamics that aim to scrutinize the causal effects and influences by analyzing the statistics of the magnitude increments and the inter-event times of stochastic processes. We derive a framework that enables to incorporate knowledge about the causal dynamics of the magnitude increments and the inter-event times of stochastic processes into a multi-fractional order nonlinear partial differential equation for the probability to find the system in a specific state at one time. Rather than following the current trends in nonlinear system modeling which postulate specific mathematical expressions, this mathematical framework enables us to connect the microscopic dependencies between the magnitude increments and the inter-event times of one stochastic process to other processes and justify the degree of nonlinearity. In addition, the newly presented formalism allows to investigate appropriateness of using multi-fractional order dynamical models for various complex system which was overlooked in the literature. We run extensive experiments on several sets of physiological processes and demonstrate that the derived mathematical models offer superior accuracy over state of the art techniques.

CCS CONCEPTS

•**Mathematics of computing** → **Probabilistic inference problems; Stochastic processes; Multivariate statistics; Information theory**; •**Computing methodologies** → **Modeling methodologies**; •**Computer systems organization** → **Embedded and cyber-physical systems**; •**Theory of computation** → **Models of learning**;

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

ICCPs, Pittsburgh, PA USA

© 2017 ACM. 978-1-4503-4965-9/17/04...\$15.00

DOI: <http://dx.doi.org/10.1145/3055004.3055017>

KEYWORDS

Complex dynamics, Causal inference, Spatio-temporal coupling, Non-linearity, Fractal dynamics

ACM Reference format:

Yuankun Xue and Paul Bogdan. 2017. Constructing Compact Causal Mathematical Models for Complex Dynamics. In *Proceedings of The 8th ACM/IEEE International Conference on Cyber-Physical Systems, Pittsburgh, PA USA, April 2017 (ICCPs)*, 11 pages.
DOI: <http://dx.doi.org/10.1145/3055004.3055017>

1 INTRODUCTION AND RELATED WORK

The recent advance in learning theory and computing power drives the rapid development of the autonomous intelligent physical system (IPS) that are cognizant and knowledge-rich. Autonomous IPS exhibits high-level awareness beyond primitive actions and employs a variety of reliable reasoning mechanisms to facilitate a decision making process under the physical systems of rich uncertainties. To enable an IPS that is able to operate in the physical world for extended periods of time with minimal or no supervision by human operators, a good understanding of the underlying complex behaviors of the physical systems is necessary.

The behavior of complex systems is influenced over many *space* and *time scales* by multi-physics interactions. From cellular interactions within the microbiome-to-brain architecture, to organ interdependencies within human body expressed via signature physiological processes, to animal swarms and food webs, to social groups and society, **memory**, **interdependency** and **concurrency** are fundamental characteristics. Although recent advances in sensing technology contributed to large datasets, the modeling and analysis of complex systems have mainly focused on modeling frameworks that ignore the exhibited long-range memory, spatio-temporal fractality, non-linear, non-ergodic and non-Gaussian properties. More precisely, traditional mathematical modeling approaches in nonlinear dynamical systems, statistical machine learning, statistical signal processing and process control, system identification and econometrics focussed on describing the complex system dynamics via a set of integer-order ordinary differential equations (ODEs) (see eq. (1)) and identify the parameters of the model by minimizing a cost function that depends on a goodness-of-fit metric (i.e., the difference between the data and the postulated model class) and a complexity of the model penalty metric [12]:

$$\frac{dx_j(t)}{dt} = f_j(x_1, \dots, x_n, u_1, \dots, u_m, \theta_j) \quad (1)$$

where the function f_j encodes the linear/nonlinear interdependencies between the state variables x_1, \dots, x_n and the control variables

u_1, \dots, u_m , and θ_j represents the set of parameters to be estimated from measured time series.

For instance, with the aim of determining equations of motion from observations of time-dependent behavior, Crutchfield and McNamara developed an informational measure to quantify the modeling accuracy for eq. (1) from time series observations and identifies the parameters of the ODEs by minimizing the distance between the data and the postulated model [6]. Relying on the assumption that the observed dynamical system is memoryless (i.e., the rate of change of its state variables obeys an integer order time derivative), the heuristic modeling strategy proposed in [18] first constructs a phase-space plot from the observations of a single variable (ignoring the true causal interdependency) and determines the dimensionality of the system's attractor. Building on postulating a particular integer order time derivative coupled with a specific nonlinear mathematical expression and using the Ritt's algorithm of differential algebra, one can find the parameters of the model by solving a regression problem [11]. Using a bilinear approximation of the dynamics of interactions between system variables, the dynamic causal modeling in [10] proposes a Bayesian framework that determines the parameters of a multi-dimensional linear (integer order) dynamical system (the dynamics is assumed to be memoryless). Along the same lines, numerous other approaches (e.g., manifold learning [15], Bayesian networks [1], Q-learning [13], variational inference [17]) have been proposed in the literature. More recently, a kernel cross-spectral density based analysis of stationary time series was proposed for measuring the independence and the similarity between various time series [3].

While not being able to be exhaustive in covering all previous work, the implicit assumption in all these prior approaches is that the *dynamics is intrinsically governed by a first order time derivative which implies that the inter-event times between successive changes in the magnitude of the processes are characterized by an exponential law*. However, many complex systems exhibit **long-range memory** [23] and **fractal dynamics** that are characterized by power (non-exponential) law magnitude and inter-event times bringing into question whether the nonlinearity should be considered in the time or space domains or both [2, 4, 22, 25].

In addition, a number of outstanding challenges remain for constructing mathematical models of complex systems: (i) How can we distinguish between spatial and temporal nonlinearity and how can we construct mathematical models (i.e., dynamical equations) that capture the spatio-temporal statistical characteristics of complex systems? (ii) How can we identify the mathematical expressions of the functions f_j for the nonlinear models of complex systems and determine the degree of nonlinearity that should be accounted for without incorporating unnecessarily many nonlinear terms? (iii) How can the power law and non-Gaussian properties of the magnitudes observed in many time series impact the degree of nonlinearity? (iii) How one can interpret the asymmetry observed in the time series realization of various processes and estimate the amount of information gained from analyzing the spatio-temporal complexity present in time series?

To overcome the afore-mentioned drawbacks, in this paper, we seek to understand mathematically how various fundamental and essential components of the complex systems interact and exchange information to influence the overall performance and behavior. To

the best of our knowledge, this paper is the first to investigate the impact of microscopic dynamics encapsulated in the ordered sequence of magnitude increments and inter-event times of the stochastic process. More precisely, **we have made the following novel contributions in this work:**

Firstly, we show that by adopting a causal inference like framework and combining with probabilistic tools that were originally developed in statistical physics context we can develop mathematical strategies that can enable us to distinguish when a time series exhibits a short-range or a long-range dependence dynamics.

Secondly, we show how the analysis of a multi-point probability density function can be interpreted through the lens of maximum entropy principle and distinguish between a memoryless or a complex time-dependency structure.

Thirdly, this analysis allows to identify conditions under which a complex system under investigation can be approximated through a single order (i.e., the inter-event times can be characterized by a marginal power law distribution being reminiscent of mono-fractal dynamics) or multiple order (i.e., the inter-event times and overall system dynamics requires multiple interwoven power laws and multiple scaling exponents being reminiscent of multi-fractal dynamics) fractional dynamics. With respect to the afore-mentioned question (i), the proposed mathematical formalism allows us to determine when the space and time components of a stochastic process can be decoupled and its overall evolution be described through a multi-fractional nonlinear partial differential equation (PDE) for the probability to reach a state at a particular point in time.

With respect to the afore-mentioned question (ii), we demonstrate how the power law behavior exhibited in the space (magnitudes) and time (intervals of time between successive changes in the magnitude) domains can be encapsulated through a compact fractional calculus representation for the probability to reach a state at a particular point in time and how this new formalism enables the derivation of the nonlinear dynamical equations with minimal postulation. Simply speaking, the statistical analysis of the magnitudes and the inter-event times dictates the mathematical expression (form) of the dynamical model.

With respect to the afore-mentioned question (iii), we illustrate how the asymmetry observed between the positive and negative increments in the magnitude of the stochastic process can be encoded through Riesz-Feller fractional order operators leading to a new class of multi-fractional nonlinear PDEs that could be exploited for computing information theoretic metrics regarding the modeling accuracy and overfitting. This is left for future work.

2 COMPACT MATHEMATICAL MODELING OF COMPLEX DYNAMICS

We show in Figure 1 the overview of the proposed framework that enables the construction of the compact mathematical causal modeling of the complex dynamics in a wide spectrum of applications in cyber-physical domain. By setting up sensors that monitor the physical process of interest, we are able to investigate the microscopic dynamics of the underlying system evolution in both space

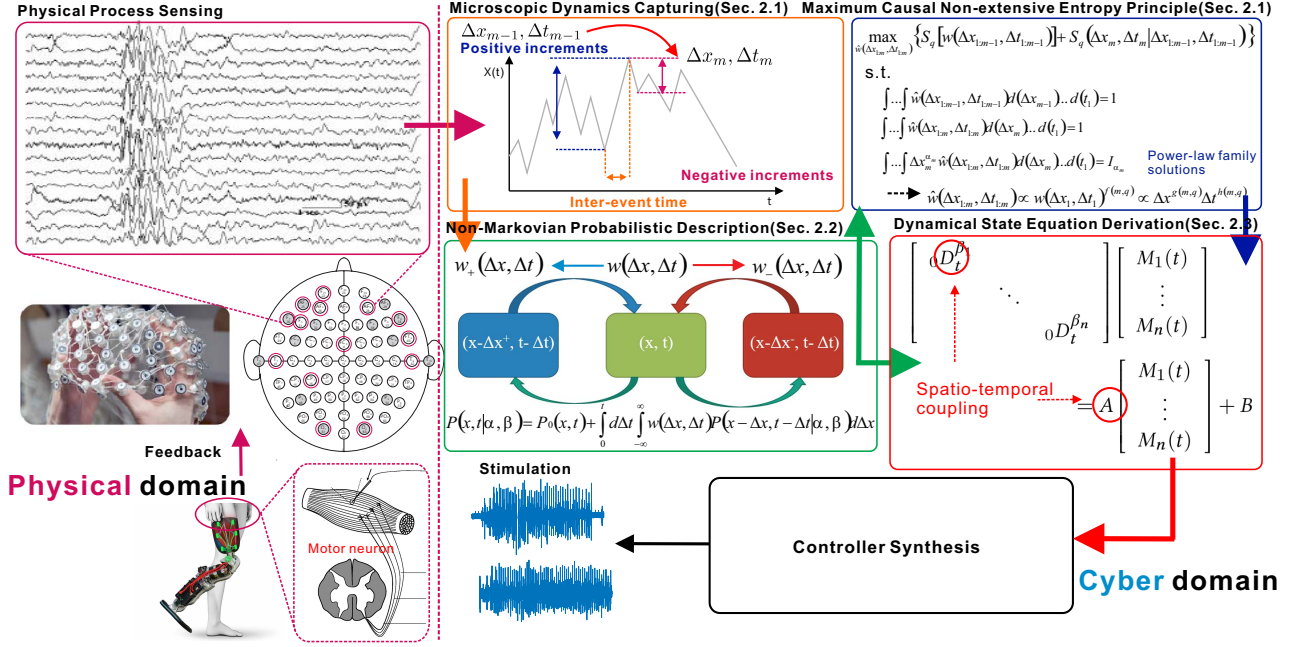


Figure 1: The overview of the proposed framework to construct compact mathematical modeling of complex dynamics and its application to the development of closed-loop cyber-physical systems. The monitored physical process is interfaced with a data-driven learning framework that investigates the system dynamics from a microscopic perspective (e.g., magnitude and inter-event time increments). A non-Markovian probabilistic description is constructed based on generalized master equation and the principle of maximum causal entropy to encapsulate the exhibited power-law behaviors (Section 2.1 and Section 2.2). This master equation based formalism allows the derivation of the nonlinear dynamical state equations with minimal postulation to capture system dynamics (Section 2.3). By retrieving the system model parameters from a specific physical process from which the model is derived, the proposed framework can serve as the intelligent core of the closed-loop intelligent cyber-physical systems for a wide spectrum of CPS applications (e.g., decode the neuron activities to generate stimulation signals for the control of prosthesis or muscles to help amputee or paralyzed patients to regain body functionality).

(magnitude increments) and time (inter-event times). We demonstrate that the interdependency between the jumps of the magnitude and inter-event times i) dictate the mathematical characterization of changing rate of system states and ii) impact the degree of nonlinearity of the couplings among system components. To capture these properties, we propose a non-Markovian probabilistic description of the physical process through generalized master equation. By principle of maximum causal entropy considering the power-law behavior exhibited in the system dynamics, we are able to derive the nonlinear dynamical state equations with minimal postulation.

By leveraging the system identification techniques to retrieve the model derived by learning the microscopic dynamics of the observed physical process, the proposed causal modeling of complex dynamics can be integrated as the core of the intelligent cyber-physical systems in a widely ranged CPS applications such as the prediction of system dynamic behaviors, the detection of system anomalies and the synthesis of the closed-loop controllers for autonomous systems. In what follows, we will present the detailed discussion on all components of our proposed mathematical modeling framework.

2.1 Microscopic Dynamics Dictates the Governing Mathematical Equation

We consider a stochastic process $x(t)$, whose realization is described by a tuple sequence of magnitude and time increments: $\{(\Delta x_1, \Delta t_1); (\Delta x_2, \Delta t_2); \dots; (\Delta x_m, \Delta t_m)\}$ (see Figure 1). More precisely, the Δt_j represents the waiting time in which the stochastic process makes a jump Δx_j in the j -th iteration. To define the underlying stochastic process (sequence), we introduce the conditional probability density function

$$w(\Delta x_m, \Delta t_m | \Delta x_{m-1}, \Delta t_{m-1} \dots; \Delta x_1, \Delta t_1) = w(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1}) \quad (2)$$

and the joint probability density function (PDF):

$$w(\Delta x_m, \Delta t_m; \dots; \Delta x_1, \Delta t_1) = w(\Delta x_{1:m}, \Delta t_{1:m}) = \prod_{j=1}^m w(\Delta t_j | \Delta t_{1:j-1}) \prod_{j=1}^m w(\Delta x_j | \Delta x_{1:j-1}, \Delta t_{1:m}) \quad (3)$$

where we have taken into account the chain rule and the causal dependency. The rationale for considering these probabilities is motivated by the need to understand how the so called microscopic dynamics influences the overall evolution; alternatively stated we

aim to understand: *i*) How the changes in the magnitude increments affect the degree of linearity /nonlinearity of the function f_j in equation (1); *ii*) How the statistics of the inter-event times impact the overall dynamics and could dictate the mathematical operator characterizing the rate of change in the system.

One fundamental assumption we make is that the tuple $(\Delta x_j, \Delta t_j)$ is not random and independent of the previous magnitude increments $(\Delta x_i, \Delta t_i)$ with $1 \leq i < j$. Instead, we assume that the ordered sequence $\{(\Delta x_1, \Delta t_1); (\Delta x_2, \Delta t_2); \dots; (\Delta x_m, \Delta t_m)\}$ possesses some form of directed information that is conveyed either from the space of magnitudes Δx_j to that of time increments Δt_j or from the space of time increments Δt_j to that of the process magnitudes Δx_j from one iteration (generation) to another (preserving the axis of time). This ordered dependency represents the source of the fractal behavior that could be observed in either the magnitudes or the waiting (inter-event) times. Consequently, in this work, we investigate how this directional information and its related statistics determine the type of dynamical equation that governs the evolution of process $x(t)$. An important problem in establishing a governing dynamical state equation (i.e., a differential equation describing the evolution of the k -th order statistical moments) is represented by the need to elucidate the dependency structure between the process magnitudes and time increments and the mathematical form of the joint PDF $w(\Delta x_m, \Delta t_m; \dots; \Delta x_1, \Delta t_1)$.

In order to describe a strategy for investigating the dependency structure between the process magnitudes and time increments, we assume that the stochastic process $x(t)$ corresponds to a non-extensive¹ system and introduce the following definitions:

Definition 1: ([24]) Given a continuous probability distribution $f(x, y)$, the non-extensive Tsallis entropy is defined by:

$$S_q[f] = \frac{1}{q-1} \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \{f(x, y) - f(x, y)^q\} dx dy \quad (4)$$

where q is a real number and represents the entropic index. When $q \rightarrow 1$ the above formula reduces to the Boltzmann-Gibbs or Shannon entropy (up to a constant parameter).

Definition 2: Given two continuous joint probability distributions $w(\Delta x_{1:m}, \Delta t_{1:m})$ and $w(\Delta x_{1:m-1}, \Delta t_{1:m-1})$, we define the causal non-extensive Tsallis entropy as follows:

$$S_q(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1}) = \frac{1}{q-1} \int \dots \int \left\{ \frac{w(\Delta x_{1:m}, \Delta t_{1:m})}{w(\Delta x_{1:m-1}, \Delta t_{1:m-1})} \left[\frac{w(\Delta x_{1:m}, \Delta t_{1:m})^{1-q}}{w(\Delta x_{1:m-1}, \Delta t_{1:m-1})^{1-q}} - 1 \right] \right\} \quad (5)$$

As will be shown in the next subsection, the mathematical expression of the joint PDF $w(\Delta x_{1:m}, \Delta t_{1:m})$ plays a crucial role in establishing the dynamical equations of the k -th order statistical moments. Consequently, in order to elucidate the form of the joint PDF $w(\Delta x_{1:m}, \Delta t_{1:m})$ we employ the principle of maximum entropy which describes a probability distribution estimator that best represents the state of knowledge or known properties of that distribution [14]. Simply speaking, this estimator implies maximizing an entropic functional subject to constraints that reflect

our knowledge about the distribution (e.g., normalization condition, fractional order statistical moments). In what follows, we will describe a maximum entropy inspired estimator for the joint PDF $w(\Delta x_{1:m}, \Delta t_{1:m})$ and its related probabilistic components.

Definition 3: The *principle of maximum causal non-extensive entropy* describes the causal non-extensive entropy-maximizing probability distribution estimator, $\hat{w}(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1})$, by solving the following optimization problem:

$$\max_{\hat{w}(\Delta x_{1:m}, \Delta t_{1:m})} \{S_q[\hat{w}(\Delta x_{1:m-1}, \Delta t_{1:m-1})] + S_q(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1})\} \quad (6)$$

$$\int \dots \int \hat{w}(\Delta x_{1:m}, \Delta t_{1:m}) d(\Delta x_m) \dots d(\Delta t_1) = 1 \quad (7)$$

$$\int \dots \int \hat{w}(\Delta x_{1:m-1}, \Delta t_{1:m-1}) d(\Delta x_{m-1}) \dots d(\Delta t_1) = 1 \quad (8)$$

$$\int \dots \int \Delta x_m^{\alpha_{m-1}} \hat{w}(\Delta x_{1:m-1}, \Delta t_{1:m-1}) d(\Delta x_{m-1}) \dots d(\Delta t_1) = I_{\alpha_{m-1}} \quad (9)$$

$$\int \dots \int \Delta x_m^{\alpha_m} \hat{w}(\Delta x_{1:m}, \Delta t_{1:m}) d(\Delta x_m) \dots d(\Delta t_1) = I_{\alpha_m} \quad (10)$$

Theorem 1: The causal non-extensive entropy-maximizing probability distribution estimator in optimization problem (6) leads to a solution for the conditional probability $\hat{w}(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1}) = \hat{w}(\Delta x_{1:m}, \Delta t_{1:m}) / \hat{w}(\Delta x_{1:m-1}, \Delta t_{1:m-1})$, where the joint PDFs $\hat{w}(\Delta x_{1:m}, \Delta t_{1:m}) = \hat{w}_m$ and $\hat{w}(\Delta x_{1:m-1}, \Delta t_{1:m-1}) = \hat{w}_{m-1}$ satisfy the following relations:

$$\hat{w}_m = \hat{w}_{m-1} h(\Delta x_{m-1:m}, \Delta t_m), \text{ where} \quad (11)$$

$$h(\Delta x_{m-1:m}, \Delta t_m) = \{(\lambda_m - \gamma_m \Delta x_m^{\alpha_m}) l(\Delta x_{m-1})^{\frac{q-2}{1-q}}\}^{\frac{1}{q-1}}$$

with $l(\Delta x_{m-1}) = 1 + (q-1)(\lambda_{m-1} + \gamma_{m-1} \Delta x_{m-1}^{\alpha_{m-1}})$

One can notice from eq. (11), that if we do not have knowledge about the fractional order statistical moments, then $w_m = w(\Delta x_{1:m}, \Delta t_{1:m})$ is proportional to the initial joint PDF $w_1 = w(\Delta x_1, \Delta t_1)$ (i.e., $w(\Delta x_{1:m}, \Delta t_{1:m}) \propto w(\Delta x_{1:m-1}, \Delta t_{1:m-1}) \dots \propto w(\Delta x_1, \Delta t_1)$). It should be noted that $w(\Delta x_m, \Delta t_m | \Delta x_{1:m-1}, \Delta t_{1:m-1})$ connects to the concept of Neyman-Rubin causality [19] and preserves the causal dependence in space and time.

2.2 Non-Markovian Probabilistic Description

In the previous subsection, we observed that under some knowledge about the stochastic process magnitudes (Δx and time increments (Δt)) reflected as constraints in the maximum causal non-extensive entropy formulation, the joint distribution can take exponential, power law or more complex mathematical expressions. In this section, we incorporate the results derived in subsection 2.1 and provide a probabilistic description of the stochastic process $x(t)$ through a generalized master equation (GME) for the conditional PDF $P(x, t | \alpha, \beta)$:

$$P(x, t | \alpha, \beta) = P_0(x, t | \alpha, \beta) + \int_0^t d(\Delta t) \quad (12)$$

$$\int_{-\infty}^{\infty} w(\Delta x, \Delta t | \alpha, \beta) P(x - \Delta x, t - \Delta t | \alpha, \beta) d(\Delta x)$$

where the $P(x, t | \alpha, \beta)$ represents the conditional PDF that the stochastic process $x(t)$ attains value x at time t given that the process magnitude increments Δx and the process inter-event times Δt are characterized by fractal exponents α and β , the $w(\Delta x, \Delta t)$ represents the joint PDF of the process magnitude increments Δx and

¹Systems obey non-extensive statistical mechanics where entropy is non-additive

the inter-event times Δt that is assumed to satisfy a fractal scaling relationship characterized by the fractal exponents α and β , and $P_0(x, t|\alpha, \beta)$ is the initial condition that the stochastic process was initiated from state x at time $t = 0$.

The conditioning on the fractal exponents α and β in eq. (12) of the PDF $P(x, t|\alpha, \beta)$ is motivated by the fact that the joint PDF of the process magnitude increments and the inter-event times $w(\Delta x, \Delta t)$ can be expressed as a generalized fractional Taylor (power) series: $w(\Delta x, \Delta t) = \sum_k (A_k + B_k \Delta x^{-\alpha_k}) \Delta t^{\beta_k-1}$. Consequently,

we express the GME for the conditional PDF $P(x, t|\alpha, \beta)$ for a specific α and β , with $w(\Delta x, \Delta t|\alpha, \beta) \propto \Delta x^{-\alpha} \Delta t^{\beta-1}$ and integrate over the set of exhibited fractal exponents. Taking into account the Riemann-Liouville fractional order integral of order $\beta > 0$ for a function f over the space of locally integrable function [16]:

$$I_t^\beta f(t) = \frac{1}{\Gamma(\beta)} \int_0^t \tau^{\beta-1} f(t-\tau) d\tau \quad (13)$$

the relation between the fractional order integral and fractional order derivative (for $0 < \beta < 1$)

$$\begin{aligned} {}_0D_t^\beta f(t) &= \frac{d^\beta f(t)}{dt^\beta} = {}_0I_t^{-\beta} f(t) = \\ &= \frac{1}{\Gamma(1-\beta)} \int_0^t \frac{1}{(t-\tau)^{\beta-1}} \frac{df(\tau)}{d\tau} d\tau \end{aligned} \quad (14)$$

and the fractional Kramer-Moyal expansion, the GME (12) [8] can be rewritten as follows:

$$\begin{aligned} {}_0D_t^\beta P(x, t|\alpha, \beta) &= \int_{-\infty}^x d\psi l^*(x-\psi) [P(\psi, t|\alpha, \beta) \\ &- P(x, t|\alpha, \beta)] + \int_x^\infty l(x-\psi) [P(\psi, t|\alpha, \beta) - P(x, t|\alpha, \beta)] d\psi \end{aligned} \quad (15)$$

The transition functionals $l^*(z)$ and $l(z)$ can take many different mathematical expressions. In what follows, we consider a generalized mathematical form motivated by observed statistical asymmetry of several physiological time series as follows:

$$l^*(z) = a_1 \frac{1}{\pi} \Gamma(1+\alpha) \sin\left(\frac{(\alpha+\theta)\pi}{2}\right) z^{-(1+\alpha)} + a_2 z^{-1} \quad \text{for } z \geq 0 \quad (16)$$

$$l(z) = a_1 \frac{1}{\pi} \Gamma(1+\alpha) \sin\left(\frac{(\alpha-\theta)\pi}{2}\right) z^{-(1+\alpha)} + a_2 z^{-1} \quad \text{for } z < 0 \quad (17)$$

Although other mathematical expressions can be retrieved from statistical analysis of the magnitude increments observed in the time series, the analysis of their implications on the overall mathematical form of the master equation is left for future work. For the above power law asymmetric transition functionals, the GME in (18) takes the following form:

$${}_0D_t^\beta P(x, t|\alpha, \beta) = \frac{a_1 \partial^\alpha P(x, t|\alpha, \beta)}{\partial |x|^\alpha} + a_2 P(x, t|\alpha, \beta) \quad (18)$$

Taking into account that the magnitude increments and the inter-event times can be characterized by a distribution of fractional

exponents, then the above partial differential equation can be rewritten as follows:

$$\begin{aligned} \int_{\beta_{min}}^{\beta_{max}} d(\beta) {}_0D_t^\beta P(x, t) d\beta &= \int_{\alpha_{min}}^{\alpha_{max}} a_2 P(x, t) e(\alpha) d\alpha + \\ &+ \int_{\alpha_{min}}^{\alpha_{max}} a_1 \frac{\partial^\alpha P(x, t)}{\partial |x|^\alpha} e(\alpha) d\alpha \end{aligned} \quad (19)$$

where $d(\beta)$ denotes the distribution of fractional exponents β characteristic to the statistics of the inter-event times, β_{min} and β_{max} are the lower and upper bounds on the exhibited fractional exponents for inter-event times, $e(\alpha)$ is the distribution of fractional exponents α corresponding to the statistics of the magnitude increments of $x(t)$, and α_{min} and α_{max} are the lower and upper limits on the fractional exponents characterizing the magnitude increments.

Investigating and determining the dynamical equations governing the evolution of complex systems implies analyzing and characterizing the statistical interdependency of multi-dimensional *signature* processes. Towards this end, by assuming that the inter-event times between changes in all the *signature* processes are characterized by a fractal exponent β , a multi-dimensional generalization of eq. (15) reads:

$$\begin{aligned} P(x_1, \dots, x_n, t|\alpha_1, \dots, \alpha_n, \beta_j) &= P_{t=0}(x_1, \dots, x_n|\alpha_1, \dots, \beta) \\ &+ \int_0^t d\tau \int_{-\infty}^x d\xi_1 \dots \int_{-\infty}^x d\xi_n l^*(x_1 - \xi_1, \dots, x_n - \xi_n, t - \tau) \\ &[P(\xi_1, \dots, \xi_n, \tau|\alpha_1, \dots, \beta) - P(x_1, \dots, x_n, t|\alpha_1, \dots, \beta)] \\ &+ \int_0^t d\tau \int_x^\infty d\xi_1 \dots \int_x^\infty d\xi_n l(x_1 - \xi_1, \dots, x_n - \xi_n, t - \tau) \\ &[P(\xi_1, \dots, \xi_n, \tau|\alpha_1, \dots, \beta) - P(x_1, \dots, x_n, t|\alpha_1, \dots, \beta)] \end{aligned} \quad (20)$$

where the transition probabilities $l^*(z_1, \dots, z_n, \tau)$ and $l(z_1, \dots, z_n, \tau)$ are introduced to capture the possible asymmetry in the evolution of the positive and negative increments of each process $x_j(t)$ and the interdependency between all processes $x_1(t), \dots, x_n(t)$.

In the master equation (20), the transition probabilities l help at capturing both the *causal structure* (i.e., how one or a set of system variables influence a specific variable and with what degree / strength, how system variables are impacted by external perturbations) and the *causal dynamics* (i.e., how much of the past evolution and with what strength influences the rate of change in one or multiple system variables). Alternatively stated, the master equation encodes the spatio-temporal interdependency and memory (changes in the patterns of interactions between system variables and variations in external inputs if needed) of a complex system. In the next subsection, we will consider a particular symmetric case, when the $l^*(z_1, \dots, z_n, \tau)$ and $l(z_1, \dots, z_n, \tau)$ take the following form:

$$l(z_1, \dots, z_n, \tau) = \sum_{\substack{j=1, \\ k \neq j}}^n [a_j + b_j z_j^{-1} + c_j z_j^{-1} z_k^{-1}] \tau^{\beta_j-1} \quad (21)$$

Under these transition probabilities, the master equation (20) can be expressed as follows:

$$P(x_1, \dots, x_n, t | \beta_j) = I^{\beta_j} \sum_{\substack{j=1, \\ k \neq j}}^n [a_j P(x_1, \dots, x_n, t | \beta_j) + b_j \frac{\partial P(x_1, \dots, x_n, t | \beta_j)}{\partial x_j} + c_j \frac{\partial^2 P(x_1, \dots, x_n, t | \beta_j)}{\partial x_j \partial x_k}] \quad (22)$$

where the inter-event times are characterized by a fractal exponent β_j that is determined by the intrinsic degree of memory associated with process x_j . We also express the conditional probability because in some cases the process $x_j(t)$ can be characterized by multiple fractal exponents and so a distribution $d(\beta_j)$ is required to fully characterize its dynamics.

2.3 Dynamical State Equations

One dimensional case: In what follows, we show how the aforementioned analysis helps at determining the time dependence of the statistical moments associated with the stochastic process $x(t)$ and defined as follows:

$$M_k(t) = \mathbb{E}[x^k] = \int_{-\infty}^{\infty} |x|^k P(x, t) dx \quad (23)$$

where k denotes the order of the statistical moments and $P(x, t)$ is assumed to be described by equation (19).

By multiplying both sides of equation (19) with $|x|^k$ and integrating over the x -magnitude space, we obtain the following relation for the k -th order moment of $x(t)$:

$$\begin{aligned} \int_{\beta_{min}}^{\beta_{max}} h(\beta) \frac{\partial^\beta M_k(t)}{\partial t^\beta} d\beta &= \int_{\alpha_{min}}^{\alpha_{max}} a_2 M_k(t) e(\alpha) d\alpha + \\ &+ \int_{\alpha_{min}}^{\alpha_{max}} a_1 e(\alpha) \int_{-\infty}^{\infty} |x|^k \frac{\partial^\alpha P(x, t)}{\partial |x|^\alpha} dx d\alpha \end{aligned} \quad (24)$$

Making use of the integration by parts formula leads to

$$\begin{aligned} \int_{\beta_{min}}^{\beta_{max}} h(\beta) \frac{\partial^\beta M_k(t)}{\partial t^\beta} d\beta &= \int_{\alpha_{min}}^{\alpha_{max}} a_2 M_k(t) e(\alpha) d\alpha + \\ &\int_{\alpha_{min}}^{\alpha_{max}} a_1 e(\alpha) \int_{-\infty}^{\infty} \frac{\partial^\alpha M_k(t)}{\partial |x|^\alpha} P(x, t) dx d\alpha \end{aligned} \quad (25)$$

and using the following result from fractional calculus

$$\frac{\partial^\alpha [x^k (a + bx)^m]}{\partial x^\alpha} = \frac{a^\alpha \Gamma(k+1) x^{k-\alpha} (a + bx)^{m-\alpha}}{\Gamma(k+1-\alpha)} \quad (26)$$

equation (24) can be writted as follows:

$$\begin{aligned} \int_{\beta_{min}}^{\beta_{max}} h(\beta) \frac{\partial^\beta M_k(t)}{\partial t^\beta} d\beta &= \int_{\alpha_{min}}^{\alpha_{max}} a_2 M_k(t) e(\alpha) d\alpha + \\ &+ \int_{\alpha_{min}}^{\alpha_{max}} a_1 \frac{\Gamma(k+1)}{\Gamma(k+1-\alpha)} M_{k-\alpha}(t) e(\alpha) d\alpha \end{aligned} \quad (27)$$

Of note, for $d(\beta) = \delta(\beta - \beta_0)$, $a_1 = 0$, considering noise terms and discretizing equation (27) using the Grünwald-Letnikov discrete fractional differential operator [20] we obtain the well know autoregressive fractionally integrated moving average (ARFIMA) type of models [7][9]. Furthermore, equation (27) represents a new class of mathematical models that could be used to study time series.

Multi-dimensional case: Modeling complex dynamical systems and determining governing differential equations for various state variables (e.g., mean, variance, skewness, kurtosis associated with a stochastic process) requires to characterize the statistical inter-dependency between several processes. While the transition probabilities can in general take many forms, in what follows, we analyze the case in which the evolution of the magnitudes and inter-event times associated with a set of interdependent processes is described by a joint PDF of the form in (21). By multiplying both sides of equation (22) by x_j , integrating the equation over the entire space (all variables x_1 to x_n) and taking into account the inverse relationship between the fractional order integral and fractional order derivative operators, we can obtain a dynamical equation for the mean $M_j(t)$ of the process x_j :

$${}_0 D_t^{\beta_j} M_j(t) = \sum_{\substack{j=1, \\ j \neq k}}^n [a_j M_j(t) - b_j + c_j M_k(t)] \quad (28)$$

or in matrix format:

$$\begin{bmatrix} {}_0 D_t^{\beta_1} & & \\ & \ddots & \\ & & {}_0 D_t^{\beta_n} \end{bmatrix} \begin{bmatrix} M_1(t) \\ \vdots \\ M_n(t) \end{bmatrix} = A \begin{bmatrix} M_1(t) \\ \vdots \\ M_n(t) \end{bmatrix} + B \quad (29)$$

where the matrices A and B contain the coefficients appearing in each fractional order differential equation for $M_j(t)$. Taking into account of Riemann-Liouville fractional order operation, one can notice the embedded capability of the proposed fractal dynamical state equation (FDSE) in capturing the inter-dependency and memory effect by the construction of master equation (19). The majority of prior works implicitly assume the dynamics of the system is governed by the first order derivative that suggests the successive changes (positive and negative increments) in magnitude of the processes are characterized by an exponential law. In contrast, equation (29) considers systems where the magnitude of the variation over time contradicts with the patterns in time-domain of commonly employed Markovian processes (e.g., Poisson approaches). The dynamics of such systems usually exhibit non-exponential and fractal behaviors characterized by a power-law.

3 EXPERIMENTS

3.1 Experiment setup

To make the discussion more concrete and exemplify the mathematical formalism, we investigated the statistical properties of a set of primary physiological processes (i.e, muscular, cardiac and neural processes). More specifically, we analyze the intramuscular EMG (iEMG), EEG and ECG signals from realistic clinical experiments.

The iEMG signals are collected at different sites of the forearm muscles as shown in Figure 2: (i) extensor digitorum (ED); (ii) flexor digitorum profundus (FDP); (iii) abductor pollicis longus (APL);

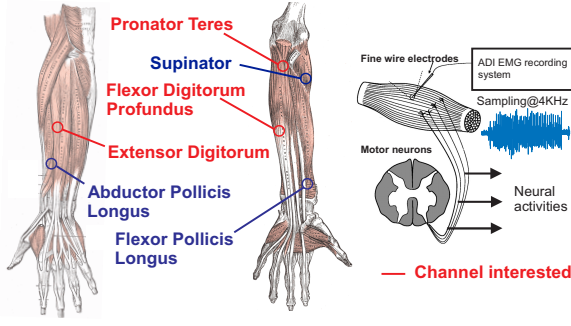


Figure 2: EMG setup: Intramuscular EMG signals are measured at 6 muscles (i.e., 2 flexor muscles, 2 extensor muscles, 1 pronator muscle and 1 supinator muscle). The subject is inserted with fine wire electrodes for measurement purpose. All channels are considered and the channels highlighted in red are selected as case study.

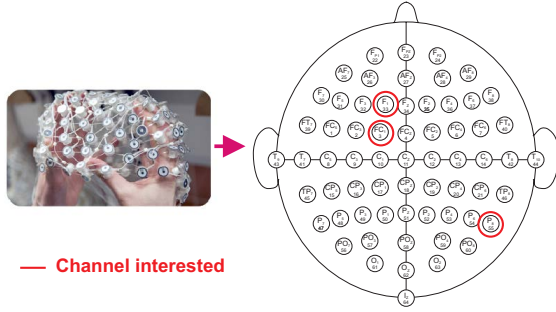


Figure 3: The 64-channel geodesic sensor distribution for measurement of EEG.

(iv) flexor pollicis longus (FPL); (v) pronator teres (PT) and (vi) supinator (SUP), when the subject is asked to relax 6 seconds, then do the finger flexion at a consistent strength for 10 seconds. The ADInstruments data acquisition system sampled the iEMG at 4 KHz.

The EEG signals are recorded by a 64-channel electroencephalogram acquisition system shown in Figure 3 that monitors the brain activity of 109 subjects when they are performing motor and imagery tasks upon noticing objects appearing on the screen [21]. Each subject is asked to open and close the corresponding fists or feet as a function of where the target appears. Each individual performed 14 experimental runs consisting of one minute with eyes open, one minute with eyes closed, and three two-minute runs of interacting with the target. The data set is collected by BCI2000 system with a sampling rate of 160Hz. The raw clinical ECG data was extracted from the PTB diagnostic ECG database [5]. Data on 52 healthy subjects (13 women, age 48 ± 19 and 39 men, age 42 ± 14) was obtained by the National Metrology Institute of Germany. Each subject's record includes 15 different signals simultaneously acquired: the conventional 12-lead (I, II, III, aVR, aVL, aVF, V1, V2, V3, V4, V5, and V6) and 3 Frank orthogonal leads (VX, VY, and VZ) as shown in Figure 4. Each signal is digitalized at 1000Hz, with a signal bandwidth of 0Hz to 1KHz and with 1 uV LSB resolution.

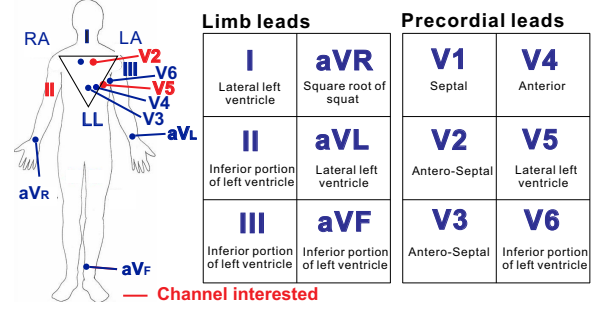


Figure 4: The deployment of 12-lead ECG system used in the experiment is shown.

3.2 Investigation of non-exponential system dynamics

In the first set of experiments, we study the stochastic nature of magnitude increments (positive and negative) in all physiological processes considered to distinguish between exponential and power law behavior or between Gaussian and non-Gaussian cases. More specifically, we perform a statistical analysis to first estimate the empirical cumulative distribution function (CDF) and its support of the physiological process magnitude positive and the absolute value of negative increments. By fitting the measurements to the postulated stochastic models, we estimate the parameters of i) an α -stable distribution, ii) exponential distribution and iii) Pareto distribution via maximum likelihood method. Then the CDFs of three stochastic models are generated based on the estimates of the model parameters.

To quantify the statistical confidence of model fitting, we performed the two-sample Kolmogorov-Smirnov test between the measurements and the generated stochastic processes with identified α -stable, exponential and Pareto distribution parameters. The null hypothesis assumes the measurements come from the same distribution as the postulated distributions with significance of 0.05. As a case study, we report the results of a selected set of physiological channels from the subjects that best illustrate their stochastic natures. Both positive and negative increments are studied and we visualize the experiments considering the positive increments in Figure 5. Figure 5.(a-c) show how two models fit to the empirical survival function (SF) of a set of EEG (a), ECG (b) and iEMG (c) signal channels in the corresponding experiments, respectively. The interested channels are highlighted in Figure 2-4. In all sub-figures, the red squares correspond to the empirical SF, the blue lines, the black lines and the green lines represent the best-fitted α -stable distribution, exponential distribution and Pareto distribution, respectively. The retrieved model parameters and p-values are reported in both the plot legends and Table 1.

By examining the figures, we can make several important observations: i) The null hypothesis that the positive increments follow an exponential distribution is rejected in all physiological channels we considered with p-value ranging from 0.001 to 0.043. This suggests that magnitude variation over time domain strongly contradicts with an exponential law which is a well-adopted assumption in previous work.

ii) The p-value well coincides with our visual inspection that the exponential SF fitting (black lines) deviates the empirical SF in all

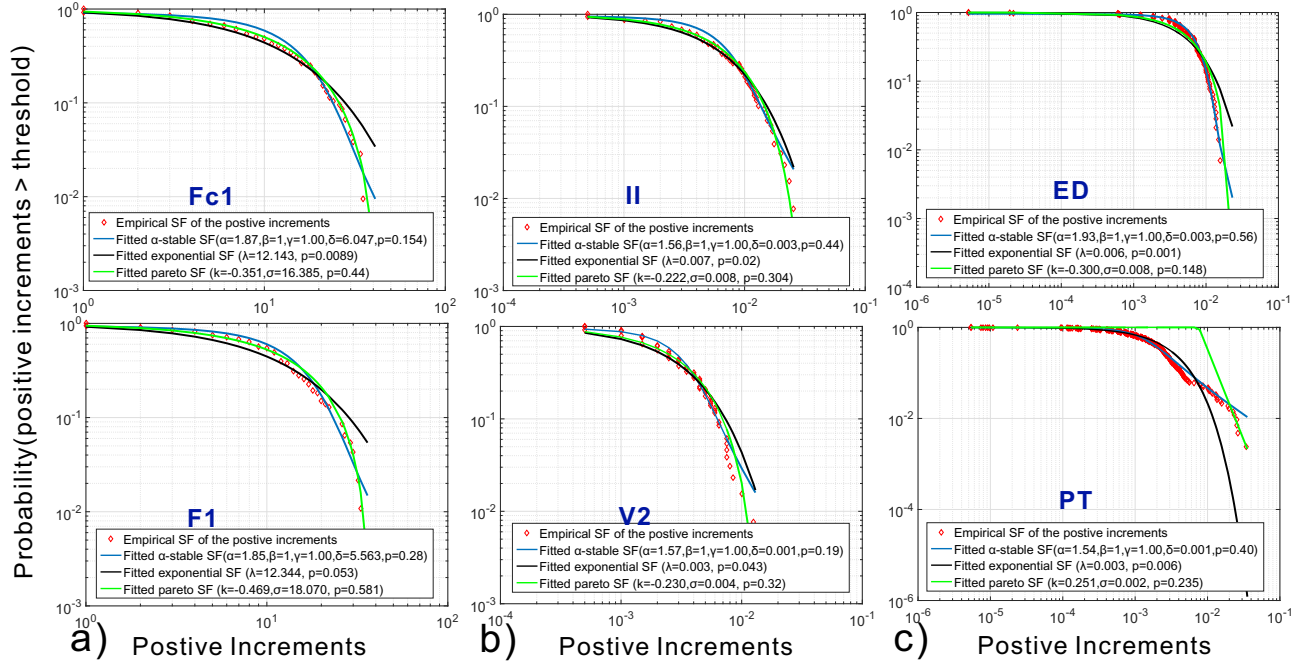


Figure 5: Empirical survival cumulative distribution functions and maximum likelihood best fitting α -stable, exponential and Pareto distributions for magnitude increments in selected EEG, ECG and EMG channels. Figure (a)-(c) show the probabilities of the positive increments in magnitude exceeding a threshold value. P-value is obtained by performing a two-sample Kolmogorov-Smirnov test with the null hypothesis that the measurements come from the same distribution as the postulated α -stable, exponential or Pareto distribution.

the channels considered. Instead, the Pareto distribution and the α -stable fitting better represent the stochastic properties of the signal variations over time in all channels. This suggests the existence of fractality which is governed by a power-law distribution as postulated by equation (21), which also coincides with our following observation.

iii) For all channels that can be better characterized by an α -stable distribution or Pareto distribution, the estimated stability parameters α are all smaller than 2 where $\alpha = 2$ corresponds to a Gaussian process. For $\alpha < 2$, stable distributions have one tail (when $\alpha < 1$ and $\beta = \pm 1$) or both tails (all other cases) that are asymptotically *power laws* with heavy tails. This is best illustrated in FDP and PT channel of iEMG signals where the empirical distribution fits well to the α -stable distribution up to a certain transition point where the empirical SF starts to obey the Pareto distribution (i.e., power-law). This suggests the existence of fractality in these physiological processes which is aligned with our analytical prediction made by Theorem 1, hence justifying the application of the proposed fractal dynamical state equation described in equation (29). The similar conclusion can also be reached by investigating the stochastic properties of the negative increments (see Table 1) in physiological processes.

3.3 Efficacy evaluation of FDSE for physiological processes

The investigation of the stochastic characteristics of the processes considered verified the existence of non-Gaussian and fractality

(non-linearity) in the physiological systems. Therefore, the dynamical behaviors of these systems in time and spatial domain can not be accurately described by the conventional methods that assumes a stationary and Markovian system state equation with short-term memory. In what follows, we present the second set of experiments where we are interested in the spatial (i.e., interdependency across multiple channels) and temporal (i.e., how previous system state passes down its influence to current system dynamics) dependency structure of the physiological processes. We evaluated the capability of the proposed FDSE in capturing the complex dynamical behaviors of physiological processes. More precisely, we employed a least-square error estimator proposed in [26, 27] to identify the model described by equation (29). After the identification of the model, we evaluate the model adequacy by comparing the physiological measurements and predicted model output as goodness-of-fit metrics.

To compare with the Vector ARMA (VARMA) model and understand the significance of coupling the fractal exponents into FDSE (i.e., considering the long term memory in system state dynamics), we report three experimental settings: i) only fractal exponents β are considered (i.e., assuming an identity matrix for A in equation (29)); ii) only coupling matrix A is considered (i.e., assuming $\beta_i = 1$ which reduces to VARMA type model) and iii) both coupling matrix and fractal exponents β are considered (i.e., FDSE). We show the comparison results of a selected set of channels from EEG (a), ECG (b) and iEMG (c) measurements in Figure 6, respectively. The blue lines show the actual measurements. The orange lines represent the predicted model output where only fractal exponents are considered (i.e., setting i). The yellow lines and the magenta lines

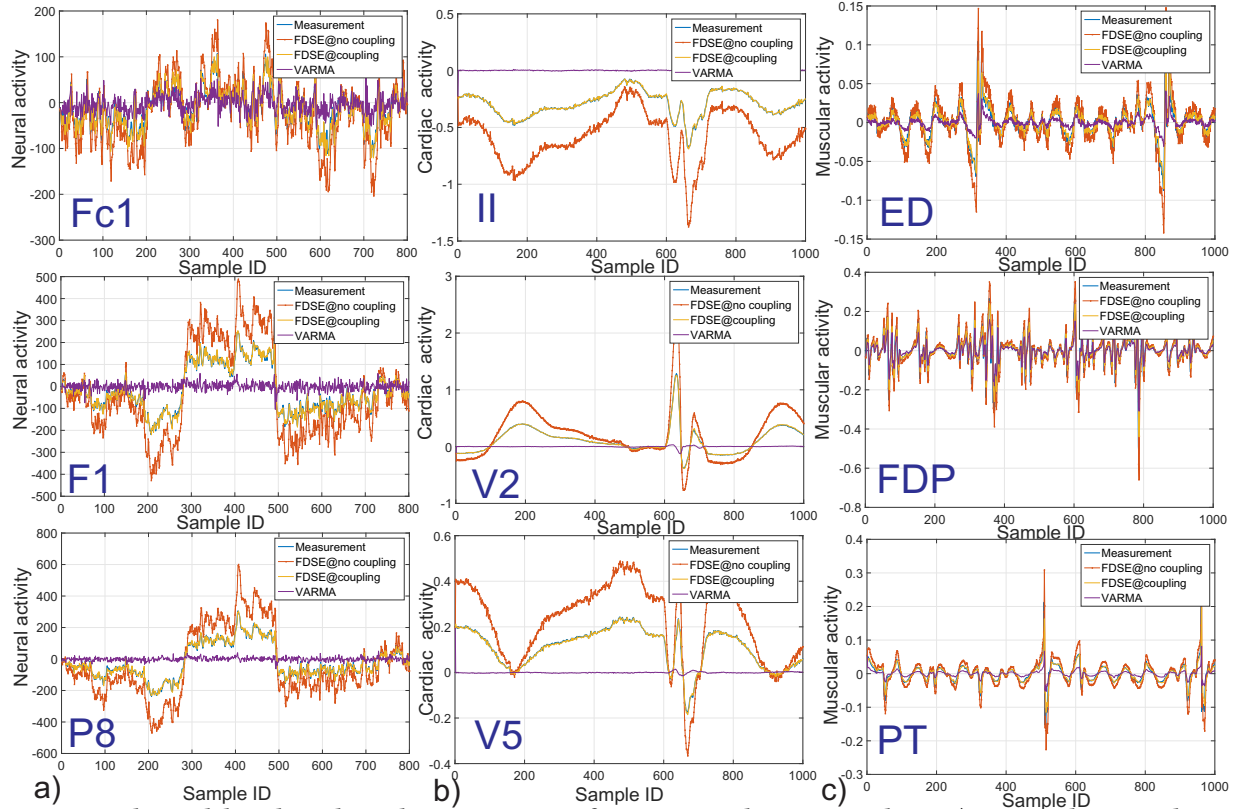


Figure 6: Fitting the model to physiological measurements of EEG, ECG and iEMG considering i) FDSE with no coupling matrix A , ii) FDSE with coupling matrix A and iii) Vector ARMA model with no fractal exponents.

Table 1: KS-test and ML best-fitting parameters of α -stable, exponential and Pareto distributions for selected channels

Channel	Positive increments						Negative increments					
	$\alpha, \beta, \gamma, \delta$	p_{stbl}	λ	p_{exp}	k, σ	p_{par}	$\alpha, \beta, \gamma, \delta$	p_{stbl}	λ	p_{exp}	k, σ	p_{par}
Fc1	1.87,1,1,6.05	0.15	12.14	0.0089	-0.35,16.39	0.44	1.56,1,1,5.56	0.08	14.47	0.071	0.39, 20.04	0.18
F1	1.85,1,1,5.56	0.28	12.34	0.053	-0.47,18.07	0.58	1.89,1,1,6.02	0.15	11.29	0.05	-0.43,16.18	0.73
P8	1.57,1,1,4.62	0.19	11.38	0.0025	-0.19, 13.56	0.16	1.88,1,1,6.17	0.01	11.2	0.03	-0.25,14.03	0.16
II	1.56,1,1,0.003	0.44	0.007	0.02	-0.22, 0.008	0.30	1.69,1,1,0.003	0.24	0.006	0.0089	-0.20, 0.007	0.21
V2	1.57,1,1,0.001	0.19	0.003	0.043	-0.23,0.004	0.32	1.8,1,1,0.002	0.05	0.004	0.01	-0.40,0.005	0.38
V5	1.4,1,1,0.001	0.15	0.003	0.0017	-0.19,0.004	0.33	1.93,1,1,0.002	0.01	0.003	0.047	-0.42,0.005	0.49
ED	1.93,1,1,0.003	0.56	0.006	0.001	-0.30,0.008	0.15	1.92,1,1,0.003	0.1	0.005	0.034	-0.29,0.007	0.65
FDP	1.09,1,0.59,0.004	0.36	0.014	0.01	0.18,0.011	0.22	0.49,1,1,0.003	0.15	0.021	0.0001	0.87,0.007	0.47
PT	1.54,1,1,0.001	0.4	0.003	0.006	0.25,0.002	0.24	1.41,1,1,0.001	0.37	0.002	0.004	0.051,0.002	0.68

correspond to setting ii and iii, respectively. We use the system identification approach proposed in [27]. The estimated fractal exponents range from 0.30 to 0.66, 0.94 to 1.19 and 0.18 to 0.61 for EEG, ECG and iEMG signals, respectively, hence verifying the existence of fractality.

Two key observations can be made for Figure 6: i) In all experiments we considered, the efficacy of incorporating the fractal exponents that captures the long-term memory effect in system dynamics can be best illustrated by the comparison between the predicted model output and actual measurements. In spite of the difference in magnitude, the predicted model output stays close to the actual measurement in terms of preserving critical system state transition behaviors. Intuitively, the predicted model output

preserves turning points and the envelope of the state dynamics of actual physiological processes. This is primarily important for construction of a time-series model capable of characterizing the physiological processes. Vital changes in bio-markers of the physiological system usually correspond to *infrequent anomalies* (e.g., the abrupt decrease/increase in blood glucose/blood pressure or excessive brain activity caused by epilepsy). The failure of the model in capturing such vital changes translates to false negative errors and might lead to irreversible undesired consequences. In contrast, the fitted VARMA models without considering the long-term memory have the tendency to smooth away the sudden changes in model output in order to minimize least-square errors as a consequence of the regression process.

ii) Comparing the goodness-of-fit between FDSE models with and without considering coupling matrix A that encodes the inter-dependency across different channels leads to interesting findings. As shown in the figure, the FDSE without matrix considering A tends to overestimate the signal magnitude as a result of accumulating the influence of the previous system states over a long time course. In contrast, FDSE model with coupling matrix A in all experiments aligns well with the actual measurements suggesting its adequacy in characterizing the physiological processes. The performance difference can be understood as follows: the FDSE assuming an identity matrix A has no knowledge of how coupled physiological processes contribute to the state transition dynamics of each other. As a result, given a specific channel, the estimation process tries to compensate the contribution from other channels by assuming a long lasting influence from the previous system states. By incorporating the matrix A , the predicted model output is well regulated to adequately fit to actual measurements by coupling the interdependency between different channels of the physiological signals.

4 CONCLUSIONS AND FUTURE WORK

Understanding the implications of the degree of nonlinearity and the nature of fractality (i.e., distinguishing between fractality in the magnitude increments (space) of one variable with respect to other variables and the fractality in the inter-event times) represents the motivation of this work. We generalize the linear / nonlinear dynamic causal approaches, by adopting a statistical physics inspired probabilistic description of various processes representing the evolution of a complex system and incorporating the statistics of the magnitude increments and the inter-event times into the mathematical expression of the master equation.

First, this new approach allows to capture the power law and nonlinear interactions that exist between the magnitude increments and the inter-event times of one stochastic process on one hand, and the inter-dependencies between the magnitude increments and the inter-event times of one process and other processes, on the other hand. Second, it provides a mathematical strategy for modeling of complex systems whose dynamics exhibits a mixture of Markovian and non-Markovian evolution. Relying on conditional probabilistic description on how ordered sequence of magnitude increments and the inter-event times affect the overall system dynamics allows to define new multivariate causal inference techniques that take into account the non-Markovian nature of the dynamics. This is left for future work. Moreover, it allows to mathematically justify the adoption of a class of mathematical model that could potentially complement current Bayesian model selection strategies. The presented mathematical framework could be enriched by combining it with other techniques from statistical machine learning and signal processing for developing new modeling strategies for complex interdependent networks. Third, the proposed causal modeling of complex dynamics can be integrated as the core of the cognizant cyber-physical systems in a widely ranged CPS applications to be able to understand, describe, predict and control the underlying physical processes.

ACKNOWLEDGMENTS

The authors are thankful to the reviewers for their valuable feedback. The authors gratefully acknowledge the support by National Science Foundation under CAREER Award CPS-1453860 and US Army Defense Advanced Research Projects Agency (DARPA) under Grant No. W911NF-17-1-0076.

REFERENCES

- [1] Chiuso A. Regularization and bayesian learning in dynamical systems: Past, present and future. *Annual Reviews in Control* - submitted, 20XX.
- [2] J. Bassingthwaite, L. Liebovitch, and B. West. Properties of fractal phenomena in space and time. In *Fractal physiology*. Springer, 1994.
- [3] M. Besserve, N. Logothetis, and B. Schölkopf. Statistical analysis of coupled time series with kernel cross-spectral density operators. In *NIPS*, 2013.
- [4] P. Bogdan. Mathematical modeling and control of multifractal workloads for data-center-on-a-chip optimization. In *Proceedings of the 9th International Symposium on Networks-on-Chip*, page 21. ACM, 2015.
- [5] R. Bousseljot, D. Kreisler, and A. Schnabel. Nutzung der ekg-signaldatenbank cardiodat der ptb über das internet. *Biomedizinische Technik/Biomedical Engineering*, 40(s1):317–318, 1995.
- [6] J. Crutchfield and B. McNamara. Equations of motion from a data series. *Complex systems*, 1(417–452):121, 1987.
- [7] C. Granger and R. Joyeux. Essays in econometrics. chapter An Introduction to Long-memory Time Series Models and Fractional Differencing, pages 321–337. Harvard University Press, Cambridge, MA, USA, 2001.
- [8] R. Herrmann. *Fractional Calculus: An Introduction for Physicists*. World Scientific.
- [9] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.
- [10] Friston K., Harrison L., and Penny W. Dynamic causal modelling. *NeuroImage*, 19(4):1273 – 1302, 2003.
- [11] L. Ljung and T. Glad. On global identifiability for arbitrary model parametrizations. *Automatica*, 30(2):265–276, 1994.
- [12] Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1 – 12, 2010.
- [13] P. Mehta and S. Meyn. Q-learning and pontryagin’s minimum principle. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proc. of the 48th IEEE Conference on*. IEEE, 2009.
- [14] E. Montroll and M. Shlesinger. Maximum entropy formalism, fractals, scaling phenomena, and 1/f noise: a tale of tails. *Journal of Statistical Physics*, 1983.
- [15] H. Ohlsson, J. Roll, and L. Ljung. Manifold-constrained regressors in system identification. In *Decision and Control, 2008. CDC 2008. 47th IEEE Conference on*, pages 1364–1369, Dec 2008.
- [16] K. Oldham and J. Spanier. *The fractional calculus : theory and applications of differentiation and integration to arbitrary order*. Mathematics in science and engineering. Academic Press, New York, 1974.
- [17] M. Opper and G. Sanguinetti. Variational inference for markov jump processes. In J.C. Platt, D. Koller, Y. Singer, and S.T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1105–1112. Curran Associates, Inc., 2008.
- [18] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw. Geometry from a time series. *Phys. Rev. Lett.*, 45:712–716, Sep 1980.
- [19] D. Rubin. Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American Statistical Association*, 100, 2005.
- [20] J. Sabatier, O. P. Agrawal, and J. A. Tenreiro Machado. *Advances in Fractional Calculus: Theoretical Developments and Applications in Physics and Engineering*. Springer Publishing Company, Incorporated, 1st edition, 2007.
- [21] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw. Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, 51(6):1034–1043, 2004.
- [22] M. Shlesinger, G. Zaslavsky, and J. Klafter. Strange kinetics. *Nature*, 363(6424):31–37, 1993.
- [23] A. Svenkeson, B. Glaz, S. Stanton, and B. West. Spectral decomposition of nonlinear systems with memory. *Physical Review E*, 93(2), 2016.
- [24] Constantino Tsallis. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Statist. Phys.*, 52:479–487, 1988.
- [25] B. West. *Fractal physiology and chaos in medicine*. World Scientific, 2012.
- [26] Y. Xue, S. Pequito, J. Coelho, P. Bogdan, and G. Pappas. Minimum number of sensors to ensure observability of physiological systems: A case study. In *Communication, Control, and Computing (Allerton), 2016 54th Annual Allerton Conference on*, pages 1181–1188. IEEE, 2016.
- [27] Y. Xue, S. Rodriguez, and P. Bogdan. A spatio-temporal fractal model for a cps approach to brain-machine-body interfaces. In *Design, Automation & Test in Europe Conference & Exhibition (DATE), 2016*, pages 642–647. IEEE, 2016.

5 APPENDIX

Proof to Theorem 1: Introduce Lagrange multiplier $\mathcal{L}(w_{m-1}, \lambda_m, \lambda_{m-1}, \gamma_m, \gamma_{m-1}, \phi_m)$ to eq. (6). Define $\delta_\Delta = \delta(\Delta x - \Delta x_m, \Delta t - \Delta t_m)$, $\delta_0 = \delta(\Delta x_m, \Delta t_m)$, $\delta_{-1} = \delta(\Delta x_{m-1}, \Delta t_{m-1})$ and $w = w_m/w_{m-1}$ such that we have,

$$\begin{aligned} \mathcal{L}(w_{m-1}, \lambda_m, \lambda_{m-1}, \gamma_m, \gamma_{m-1}, \phi_m) = & \int \dots \int \\ & \frac{1}{q-1} (w_m - w_{m-1}) \delta_\Delta + \frac{w}{q-1} \left(\frac{w^{1-q}}{w_m} - 1 \right) \\ & + \lambda_m (w w_{m-1} - \delta_0) + \lambda_{m-1} (w_{m-1} - \delta_\Delta - \delta_{-1}) \\ & + \gamma_m (\Delta x_m^{-\alpha_m} w w_{m-1} - I_{\alpha_m} \delta_0) \\ & + \gamma_{m-1} (\Delta x_{m-1}^{-\alpha_{m-1}} w_{m-1} \delta_\Delta - I_{\alpha_{m-1}} \delta_{-1}) \end{aligned} \quad (30)$$

Take the first order derivative of \mathcal{L} in terms of w_{m-1} , $\partial \mathcal{L} / \partial w_{m-1}$ and make it equal to 0 such that,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial w_{m-1}} = 0 = & \int \dots \int \\ & \frac{1}{q-1} (1 - q w_{m-1}^{q-1}) \delta_\Delta + \left(\frac{w}{w_{m-1}} \right)^{2-q} \\ & + \lambda_m w + \lambda_{m-1} \delta_\Delta + \gamma_m \Delta x_m^{-\alpha_m} w + \gamma_{m-1} \Delta x_{m-1}^{-\alpha_{m-1}} \delta_\Delta \end{aligned}$$

Eq.(31) holds for any choice of m . Consequently, we have

$$1 - q w_{m-1}^{q-1} = (1 - q) (\lambda_m + \gamma_{m-1} \Delta x_{m-1}^{\alpha_{m-1}}) \quad (31)$$

$$\left(\frac{w}{w_{m-1}} \right)^{2-q} = -w (\lambda_m + \gamma_m \Delta x_m^{\alpha_m}) \quad (32)$$

Define $\lambda'_m = -\lambda_m$

$$w = h(\Delta x_{m:m-1}, \Delta t_m)$$

$$= \{ (\lambda'_m - \gamma_m \Delta x_m^{\alpha_m}) l(\Delta x_{m-1})^{\frac{q-2}{1-q}} \}^{\frac{1}{q-1}} \quad (33)$$

$$\text{with } l(\Delta x_{m-1}) = 1 + (q-1)(\lambda_{m-1} + \gamma_{m-1} \Delta x_{m-1}^{\alpha_{m-1}})$$