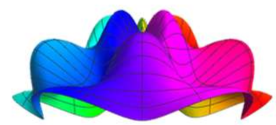
Complex-time Representation of Repeated Measurement Longitudinal Data & Space-kime Analytics

Ivo D. Dinov

https://SOCR.umich.edu



STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

1

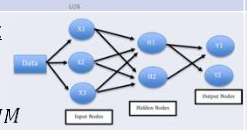
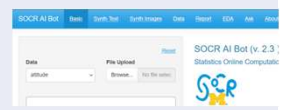
Outline

- ❑ Data Science ≡ need-based information compression & expansion
- ❑ Complex-time (*kime*) & rationale
- ❑ Kime-phase, random sampling & Heisenberg's Uncertainty
- ❑ Solutions of ultrahyperbolic wave equations
- ❑ Open spacekime problems
- ❑ Data science applications
- ❑ Bayesian formulation of spacekime inference
- ❑ Resources, live demo links & prospective DS R&D, education & practice
- ❑ *Part 2: Hands-on Spacekime Analytics Tutorial (Demos)*

2

Duality of Evidence-based Scientific Discovery

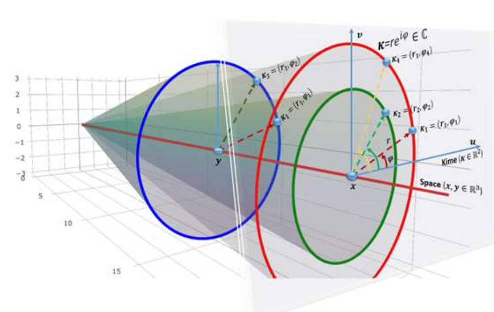
experimental → theoretical → computational → data sciences

Mapping *Examples*	**Analysis** Observables/Data → Compact Models	**Synthesis** Compact Models → (simulated, actionable info)
1. *Lossless* Math Transforms	(A.1.1) Linear transform, L: $V \to W$, e.g., *2D rigid body* $L = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}: \mathbb{R}^2 \xrightarrow{rotation} \mathbb{R}^2$ (A.1.2) Fourier transform: $\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i2\pi\omega x} dx$	(S.1.1) Inverse linear transform, $L^{-1}: W \to V$, e.g., $L^{-1} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}: \mathbb{R}^2 \xrightarrow{rotation} \mathbb{R}^2$, $LL^{-1} \equiv \mathbb{I}$ (S.1.2) Inverse Fourier (IFT): $f(x) = \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i2\pi\omega x} d\omega$
2. DNA	(A.2.1) DNA Packing in Chromatin Fiber Chromosomes contain enormously long linear DNA molecules associated with proteins that fold and pack the fine DNA double helix into a *tight compact structure*	(S.2.1) DNA Unpacking The process of unfolding the DNA from the chromosome to support the processes of gene expression, DNA replication, and DNA repair
3. *Lossy* Data/Stats Science	(A.3.1) Info Compression, e.g., linear models $Y = 4582.70 + 212.29\,X$ $Data \xrightarrow{assumptions} Model$ 	(S.3.1) Information Inflation, Simulation & Generation, e.g., forecasting, regression, interpolation, extrapolation (*predict & classify new data*): $Input \xrightarrow{mod} Output$
4. Artificial & Augmented Intelligence	(A.4.1) Building, Fitting & Training large foundational, generative & deep network AI models $Data \xrightarrow{human+infrast} GAIM$ 	(S.4.1) Generative Artificial Intelligence Modeling (GAIM) $Human\ Prompt \xrightarrow{GAIM} Result$ 

3

Complex-Time (Kime)

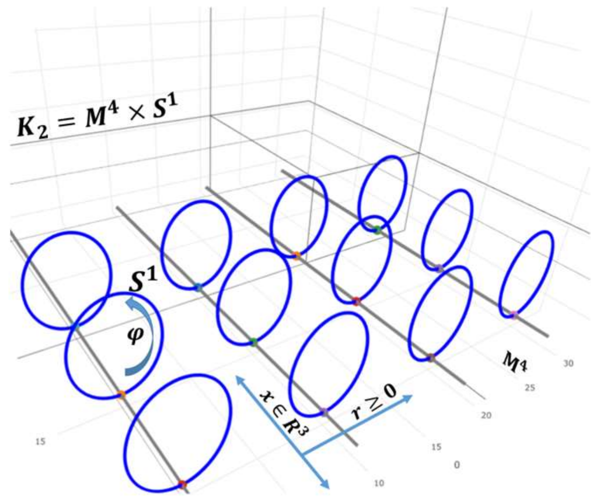
- At a given spatial location, $\boldsymbol{x}$, complex time (*kime*) is defined by $\kappa = re^{i\varphi} \in \mathbb{C}$, where:
  - the magnitude represents the longitudinal events order ($r > 0$) and characterizes the longitudinal displacement in time, and
  - event phase ($-\pi \le \varphi < \pi$) is an angular displacement, event direction, or random sampling index
- There are multiple alternative parametrizations of kime in the complex plane
- Space-kime manifold is $\mathbb{R}^3 \times \mathbb{C}$:
  - $(\boldsymbol{x}, k_1)$ and $(\boldsymbol{x}, k_4)$ have the same spacetime representation, but different spacekime coordinates,
  - $(\boldsymbol{x}, k_1)$ and $(\boldsymbol{y}, k_1)$ share the same kime, but represent different spatial locations,
  - $(\boldsymbol{x}, k_2)$ and $(\boldsymbol{x}, k_3)$ have the same spatial-locations and kime-directions, but appear sequentially in order, $r_2 < r_1$.



4

Historical Background: Kaluza-Klein Theory

- Theodor Kaluza (1921) developed a math extension of the classical general relativity theory to 5D. This included the metric, the field equations, the equations of motion, the stress-energy tensor, and the cylinder condition. Physicist Oskar Klein (1926) interpreted Kaluza's 3D+2D theory in quantum mechanical space and proposed that the fifth dimension was curled up and microscopic.
- The topology of the 5D Kaluza-Klein spacetime is $K_2 \cong M^4 \times S^1$, where M^4 is a 4D Minkowski spacetime and S^1 is a circle (non-traversable).



5

AI & Spacekime Analytics

Rationale for *Time* $\Longrightarrow$ *Kime* Extension

Math – $Time$ is a special case of *kime*, $\kappa = |\kappa|e^{i\varphi}$ where $\varphi = 0$

Time ($\mathbb{R}^+$) is a subgroup of the multiplicative Reals group

Whereas kime ($\mathbb{C}$) is an algebraically *closed prime field* that naturally extends time

Time is ordered but *kime* is not!

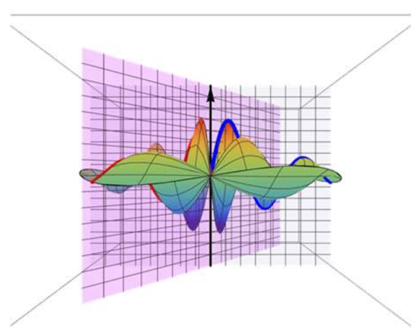
Kime ($\mathbb{C}$) represents the smallest natural extension of time, as a complete filed that agrees with time

Physics –

Problem of time … (DOI 10.1007/978-3-319-58848-3)

$\mathbb{R}$ and $\mathbb{C}$ Hilbert-space quantum theories make different predictions (DOI: 10.1038/s41586-021-04160-4)

Bio AI/Data Science – Random IID sampling, Bayesian reps, tensor modeling of $\mathbb{C}$ kimesurfaces, novel analytics



Wesson (2004, 2010)
Dinov & Velev (2021)
Wang et al. (2022)
Zhang et al. (2023)
Dinov & Shen (2024)

6

Uncertainty in 5D Spacekime

In 5D space-time, $\Omega = \left(\frac{l}{L}\right)^2$ is the conformal factor, and L is a constant length defined in terms of the cosmological constant $\Lambda = -\epsilon \frac{3}{L^2}$. In the metric signature $(+,-,-,-)$,

$\Lambda > 0$ for a spacelike extra coordinate, and $\Lambda < 0$ for a time-like extra 5th coordinate, x^μ is the $(D-1)$ spacetime location, and l is the extra kime dimension.

The *canonical spacekime metric* is:

$$dS^2 = \frac{l^2}{L^2} \sum_0^{D-2} \sum_0^{D-2} g_{\alpha\beta}(x^\mu, l) dx^\alpha dx^\beta + \epsilon dl^2$$ (5D Spacekime line element and metric)

The 4D components of the spacekime equations of motion can be written explicitly in terms of the fifth force f^μ measured in units of inertia mass, i.e., assuming $m = 1$:

$$\frac{du^\mu}{ds} + \sum_0^3 \sum_0^3 \Gamma^\mu_{\beta\gamma} u^\beta u^\gamma = f^\mu, \qquad f^\mu \equiv \sum_{\alpha=0}^3 \sum_{\beta=0}^3 \left(-g^{\mu\alpha} + \frac{1}{2} u^\mu u^\alpha\right) \frac{dl}{ds} \frac{dx^\beta}{ds} \frac{\partial g_{\alpha\beta}}{\partial l}$$

The 5D component of the spacekime equation of motion is:

$$\frac{d^2 l}{ds^2} - \frac{2}{l}\left(\frac{dl}{ds}\right)^2 - \frac{l}{L^2} = \frac{1}{2}\left[\frac{l^2}{L^2} + \left(\frac{dl}{ds}\right)^2\right] \sum_{\alpha=0}^3 \sum_{\beta=0}^3 u^\alpha u^\beta \frac{\partial g_{\alpha\beta}}{\partial l} \ \ \& \ \ f_\parallel^\mu = -\frac{1}{2} u^\mu \sum_{\alpha=0}^3 \sum_{\beta=0}^3 \left(\frac{\partial g_{\alpha\beta}}{\partial l} u^\alpha u^\beta\right) \frac{dl}{ds}$$

In 5D spacekime, geodesic motion is perturbed by an extra **5th force** $f^\mu = f_\perp^\mu + f_\parallel^\mu$, where

- $f_\perp^\mu$ is normal to the 4-velocity u_μ, similar to other conventional forces, and $f_\perp^\mu u_\mu = 0$
- $f_\parallel^\mu$ is parallel to the 4-velocity u_μ, has no analog in 4D spacetime, and $f_\parallel^\mu u_\mu \neq 0$

Wesson (2004, 2010) | Wesson & Overduin (2018) | Dinov & Velev (2021)

7

Uncertainty in 5D Spacekime

- Assuming $m = 1, c = 1$, near the foliation leaf membrane hypersurface, we have

$$\langle dp \mid dx \rangle = \sum_{\mu=0}^3 dp^\mu dx_\mu = L\left(\frac{dl}{l-l_0}\right)^2 = \frac{h}{m\,c}\left(\frac{dl}{l-l_0}\right)^2 \sim h$$

  derived from 5D Einstein deterministic field equ's $\Longrightarrow$ uncertainty principle in 4D Minkowski spacetime
- In spacetime, Heisenberg's uncertainty is due to lack of sufficient information about the 2nd kime dimension, l.
- In Minkowski 4D spacetime, the lack of kime-phase information naturally leaves one degree of freedom (**DoF**) in the system, which appears as Heisenberg's uncertainty.
- **In Bioinfo/Biostatistics, Data Science, ML/AI & longitudinal analysis, this extra DoF represents process stochasticity – random sampling from an underlying probability distribution**
- Spacekime formulation of the 4D spacetime observation of the Heisenberg's principle also supports the de Broglie-Bohm theory, which provides an explicit deterministic model of a system configuration and its corresponding wavefunction
- 4D probabilistic spacetime is a spacekime embedding with an added degrees of freedom
- Bell's theorem suggests that any deterministic hidden-variable theory, which is consistent with quantum mechanics predictions, has to be non-local. This implies the existence of instantaneous, faster than the speed of light, interactions between particles that are significantly separated in 3D space (non-local relations).

Wesson (2004, 2010) | Bell (1964) | Dinov & Velev (2021)

8

Ultrahyperbolic Wave Equation – Cauchy Initial Data

- **Nonlocal constraints** yield the existence, uniqueness & stability of local and global solutions to the ultrahyperbolic wave equation under Cauchy initial data …

$$\underbrace{\sum_{i=1}^{d_s} \partial_{x_i}^2 u \equiv \Delta_{\boldsymbol{x}} u(\boldsymbol{x},\boldsymbol{\kappa})}_{\text{spatial Laplacian}} = \underbrace{\Delta_{\boldsymbol{\kappa}} u(\boldsymbol{x},\boldsymbol{\kappa}) \equiv \sum_{i=1}^{d_t} \partial_{\kappa_i}^2 u}_{\text{temporal Laplacian}}, \quad \underbrace{\begin{vmatrix} u_o = u\left(\underbrace{\boldsymbol{x}}_{\boldsymbol{x}\in D_s}, \underbrace{0,\boldsymbol{\kappa}_{-1}}_{\boldsymbol{\kappa}\in D_t}\right) = f(\boldsymbol{x},\boldsymbol{\kappa}_{-1}) \\ u_1 = \partial_{\kappa_1} u(\boldsymbol{x},0,\boldsymbol{\kappa}_{-1}) = g(\boldsymbol{x},\boldsymbol{\kappa}_{-1}) \end{vmatrix}}_{\text{initial conditions (Cauchy Data)}}$$

where $\boldsymbol{x} = (x_1, x_2, \dots, x_{d_s}) \in \mathbb{R}^{d_s}$ and $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_{d_t}) \in \mathbb{R}^{d_t}$ are the Cartesian coordinates in the d_s space and d_t time dims.

Stable local solution over a Fourier frequency region defined by nonlocal constraints $|\boldsymbol{\xi}| \geq |\boldsymbol{\eta}_{-1}|$:

$$\hat{u}\left(\boldsymbol{\xi}, \underbrace{\kappa_1, \boldsymbol{\eta}_{-1}}_{\boldsymbol{\eta}}\right) = \cos\left(2\pi\,\kappa_1 \sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}\right) \underbrace{\hat{u}_o(\boldsymbol{\xi},\boldsymbol{\eta}_{-1})}_{c_1} + \sin\left(2\pi\,\kappa_1 \sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}\right) \underbrace{\frac{\hat{u}_1(\boldsymbol{\xi},\boldsymbol{\eta}_{-1})}{2\pi\sqrt{|\boldsymbol{\xi}|^2 - |\boldsymbol{\eta}_{-1}|^2}}}_{c_2},$$

$$\text{where } \mathcal{F}\begin{pmatrix} u_o \\ u_1 \end{pmatrix} = \begin{pmatrix} \hat{u}_o \\ \hat{u}_1 \end{pmatrix} = \begin{pmatrix} \hat{u}_o(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \\ \hat{u}_1(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \end{pmatrix} = \begin{pmatrix} \hat{u}(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \\ \partial_{\kappa_1}\hat{u}(\boldsymbol{\xi},\boldsymbol{\eta}_{-1}) \end{pmatrix}.$$

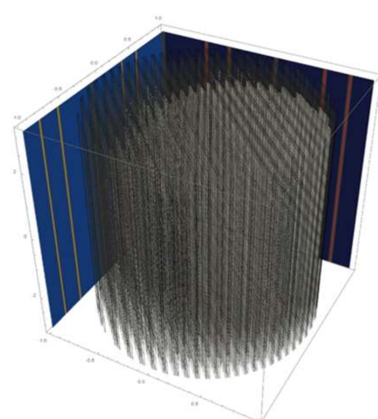
$$u\left(\boldsymbol{x}, \underbrace{\kappa_1, \boldsymbol{\kappa}_{-1}}_{\boldsymbol{\kappa}}\right) = \mathcal{F}^{-1}(\hat{u})(\boldsymbol{x},\boldsymbol{\kappa}) = \int_{\widehat{D}_s \times \widehat{D}_{t-1}} \hat{u}(\boldsymbol{\xi},\kappa_1,\boldsymbol{\eta}_{-1}) \times e^{2\pi i \langle \boldsymbol{x},\boldsymbol{\xi}\rangle} \times e^{2\pi i \langle \boldsymbol{\kappa}_{-1},\boldsymbol{\eta}_{-1}\rangle} d\boldsymbol{\xi}\, d\boldsymbol{\eta}_{-1}\ .$$

Craig & Weinstein (2008) | Wang et al. (2022) | Dinov & Velev (2021)

9

Ultrahyperbolic Wave Equation – Cauchy Initial Data

- Math Generalizations:
  Derived other spacekime concepts: law of addition of velocities, energy-momentum conservation law, stability conditions for particles moving in spacekime, conditions for nonzero rest particle mass, causal structure of spacekime, and solutions of the ultrahyperbolic wave equation under Cauchy initial data …

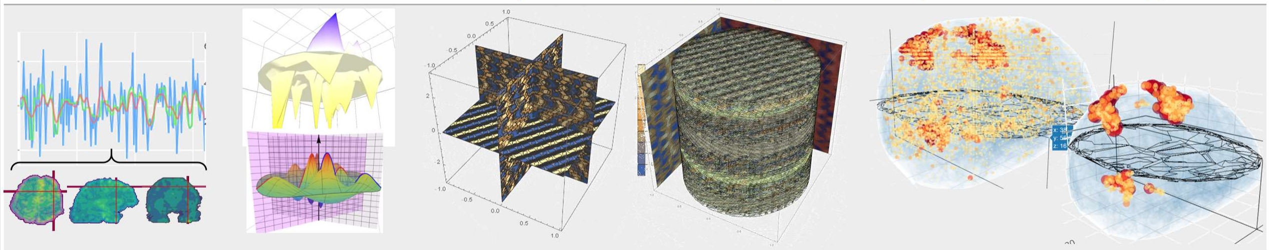


Wang et al., 2022 | Dinov & Velev (2021)

10

Longitudinal Bio Data ⟹ Kime-Transforms ⟹ PDEs ⟹ AI

Time → Kime Transformation	Wave equation Solutions (kime) dynamics	Prospective Data Science Applications



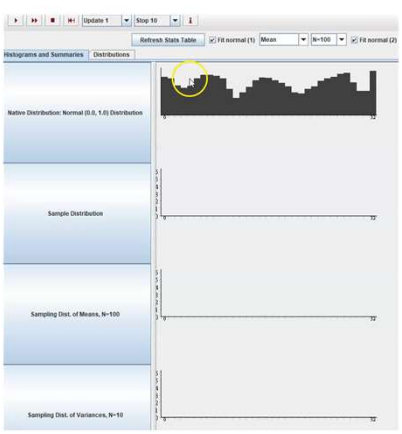
fMRI time-series   fMRI kime-surfaces   Cross sections   Volume rendering   3D p-value map   Stat significance

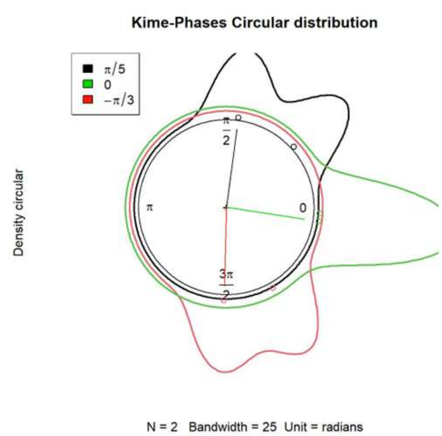
Wang et al., 2022 | Dinov & Velev (2021)

13

Random Sampling & Kime-Phase Paradigm

Kime phase distributions are mostly symmetric, random observations ≡ phase sampling





https://wiki.socr.umich.edu/index.php/SOCR_EduMaterials_Activities_GeneralCentralLimitTheorem

https://www.socr.umich.edu/TCIU/HTMLs/Chapter6_Kime_Phases_Circular.html

Dinov, Christou & Sanchez (2008)   Dinov & Velev (2021)

14

(Many) Spacekime Open Math Problems

Ergodicity

Let's look at particle velocities in the 4D Minkowski spacetime (X), a measure space where gas particles move spatially and evolve longitudinally in time. Let $\mu = \mu_x$ be a measure on X, $\underline{f(\boldsymbol{x}, t)} \in L^1(X, \mu)$ be an integrable function (e.g., velocity of a particle), and $\underline{T}: X \to X$ be a measure-preserving transformation at position $\boldsymbol{x} \in \mathbb{R}^3$ and time $t \in \mathbb{R}^+$.

A pointwise ergodic theorem argues that in a measure theoretic sense, the average of f (e.g., velocity) over all particles in the gas system at a fixed time, $\bar{f} = \mathbb{E}_t(f) = \int_{\mathbb{R}^3} f(\boldsymbol{x}, t) d\mu_x$, will be equal to the average f of just one particle ($\boldsymbol{x}$) over the entire time span,

$$\tilde{f} \equiv \mathbb{E}_x(f) = \lim_{n \to \infty} \left(\frac{1}{n} \sum_{m=0}^{n} f(T^m \boldsymbol{x}) \right), \text{ i.e., (show) } \bar{f} \equiv \tilde{f}.$$

The spatial probability measure is denoted by μ_x and the transformation $T^m \boldsymbol{x}$ represents the dynamics (time evolution) of the particle starting with an initial spatial location $T^o \boldsymbol{x} = \boldsymbol{x}$.

Investigate the ergodic properties of various transformations in the 5D **spacekime**:

$$\underbrace{\bar{f} \equiv \mathbb{E}_\kappa(f) = \frac{1}{\mu_x(X)} \int f\left(\boldsymbol{x}, \underbrace{t, \phi}_{\kappa}\right) d\mu_x}_{\text{space averaging}} \stackrel{?}{=} \underbrace{\lim_{t \to \infty} \left(\frac{1}{t} \sum_{m=0}^{t} \left(\int_{-\pi}^{+\pi} f(T^m \boldsymbol{x}, t, \phi) d\Phi \right) \right)}_{\text{kime averaging}} = \mathbb{E}_x(f) \equiv \tilde{f}$$

Dinov & Velev (2021)

15

Mathematical-Physics ⟹ Bio-Data Science & AI

Physics	**Bio-Data Sciences**
A **particle** is a small localized object that permits observations and characterization of its physical or chemical properties	An **object** is something that exists by itself, actually or potentially, concretely or abstractly, physically or incorporeal (e.g., person, subject, etc.)
An **observable** a dynamic variable about particles that can be measured	A **feature** is a dynamic variable or an attribute about an object that can be measured
Particle **state** is an observable particle characteristic (e.g., position, momentum)	**Datum** is an observed quantitative or qualitative value, an instantiation, of a feature
Particle **system** is a collection of independent particles and observable characteristics, in a closed system	**Problem**, aka Data System, is a collection of independent objects and features, without necessarily being associated with a priori hypotheses
Wave-function	**Inference-function**
Reference-Frame **transforms** (e.g., Lorentz)	Data **transformations** (e.g., wrangling, log-transform)
State of a system is an observed measurement of all particles ~ wavefunction	**Dataset (data)** is an observed instance of a set of datum elements about the problem system, $\boldsymbol{O} = \{\boldsymbol{X}, \boldsymbol{Y}\}$
A **particle system is computable** if (1) the entire system is logical, consistent, complete and (2) the unknown internal states of the system don't influence the computation (wavefunction, intervals, probabilities, etc.)	**Computable data object** is a very special representation of a dataset which allows direct application of computational processing, modeling, analytics, or inference based on the observed dataset
⋮	⋮

16

Mathematical-Physics ⟹ Bio-Data Science & AI

Physics	Data Science

Physics

<u>Wavefunction</u>

Wave equ problem:

$$\left(\frac{\partial^2}{\partial x^2} - \frac{1}{v^2}\frac{\partial^2}{\partial t}\right)\psi(x,t) = 0$$

Complex Solution:
$\psi(x,t) = Ae^{i(kx-wt)}$
represents a traveling wave,
where $\left|\frac{w}{k}\right| = v$.

Data Science

<u>Inference function</u> - describing a solution to a specific data analytic system (a problem). Examples:

- A <u>linear (GLM) model</u> represents a solution of a prediction inference problem, $Y = X\beta$, where the inference function quantifies the effects of all independent features (X) on the dependent outcome (Y), data: $O = \{X, Y\}$:

$$\psi(O) = \psi(X,Y) \;\; \Rightarrow \;\; \hat{\beta} = \hat{\beta}^{OLS} = \langle X|X\rangle^{-1}\langle X|Y\rangle = \left(X^T X\right)^{-1} X^T Y.$$

- A non-parametric, <u>non-linear</u>, alternative inference is SVM classification. If $\psi_x \in H$, is the lifting function $\psi: R^\eta \to R^d$ ($\psi: x \in R^\eta \to \tilde{x} = \psi_x \in H$), where $\eta \ll d$, the kernel $\psi_x(y) = \langle x|y\rangle: O \times O \to R$ transformes non-linear to linear separation, the observed data $O_i = \{x_i, y_i\} \in R^\eta$ are lifted to $\psi_{O_i} \in H$. The SVM prediction operator is the weighted sum of the kernel functions at ψ_{O_i}, where β^* is a solution to the SVM regularized optimization:

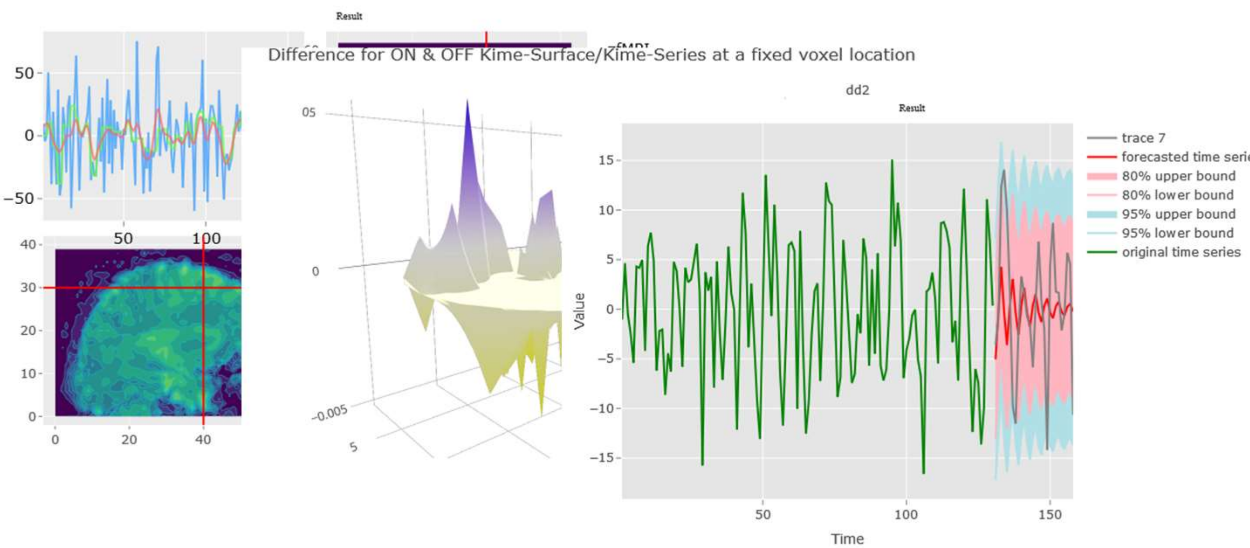
$$\underbrace{\langle \psi_O | \beta^* \rangle_H}_{predictions} = w^T x + b = \sum_{i=1}^{n} p_i^* \langle \psi_O | \psi_{O\,i} \rangle_H + b,$$

$$\min_{w \in \mathbb{R}^d,\, \xi \in \mathbb{R}^+} \left(\overbrace{\|w\|^2}^{regularizer} + C \overbrace{\sum_{i=1}^{m} \xi_i}^{fidelity} \right), y^{(i)}\left(w^T x^{(i)} + b\right) \geq 1 - \xi_i, \xi_i \geq 0$$

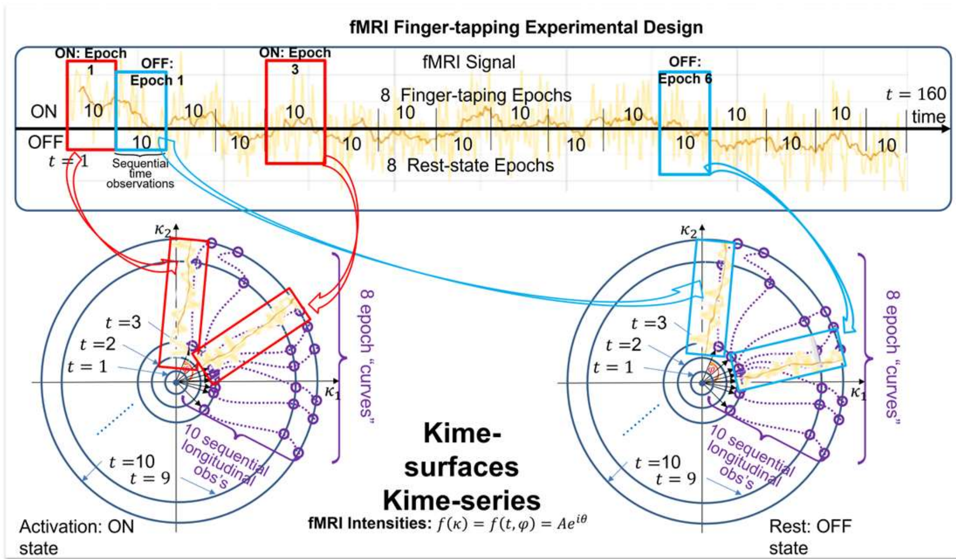
The dual weight coefficients, p_i^*, are multiplied by the label corresponding to each training instance, $\{y^{(i)}\}$.

Inference always depends on the (input) data; however, it does not have 1-1 and onto bijective correspondence with the data, since the inference function quantifies predictions probabilistically.

Spacetime Time-series ⟹ Spacekime Kimesurfaces ⟹ TLM



Mapping Longitudinal Bio-Data (Time-series) ⟹ Kime-Surfaces



Zhang et al., 2022 | Dinov & Velev (2021)

19

(Analytic) Mapping Bio Time-series ⟹ Kime-surfaces

Apply the ILT ($\mathcal{L}^{-1}$) to reconstruct a time-series, $f(t) = \mathcal{L}^{-1}(F)(t)$:

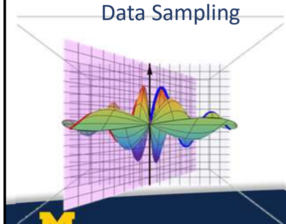
$$\boldsymbol{F(z)} = \mathcal{L}(f) = \underbrace{\frac{1}{z+1}}_{F_1(z)=\mathcal{L}(f_1(t)=e^{-t})} + \underbrace{\frac{1}{z^2+1}}_{F_2(z)=\mathcal{L}(f_2(t)=\sin(t))} \times \underbrace{\frac{z}{z^2+1}}_{F_3(z)=\mathcal{L}(f_3(t)=\cos(t))} + \underbrace{\frac{1}{z^2}}_{F_4(z)=\mathcal{L}(f_4(t)=t)} ,$$

$$\boldsymbol{f(t) = \mathcal{L}^{-1}(F)} = \mathcal{L}^{-1}(F_1 + F_2 \times F_3 + F_4)) = \mathcal{L}^{-1}(F_1) + \Big(\underbrace{\mathcal{L}^{-1}(F_2) * \mathcal{L}^{-1}(F_3)}_{\text{convolution}}\Big)(t) + \mathcal{L}^{-1}(F_4) =$$

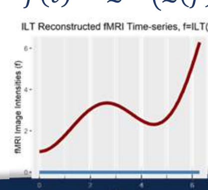
$$\mathcal{L}^{-1}(\mathcal{L}(f_1))(t) + \Big(\mathcal{L}^{-1}(\mathcal{L}(f_2)) * \mathcal{L}^{-1}(\mathcal{L}(f_3))\Big)(t) + \mathcal{L}^{-1}(\mathcal{L}(f_4))(t) ,$$

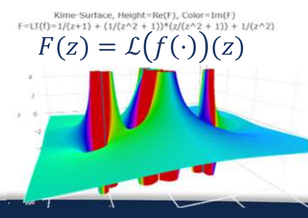
$$\boldsymbol{f(t)} = \mathcal{L}^{-1}(F)(t) = f_1(t) + (f_2 * f_3)(t) + f_4(t) = e^{-t} + \int_0^t \sin(\tau) \times \cos(t-\tau)d\tau + t = \boldsymbol{t + e^{-t} + \frac{t\sin(t)}{2}} .$$

Repeated Longitudinal Data Sampling



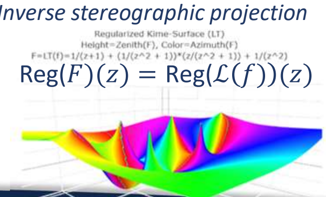
$f(t) = \mathcal{L}^{-1}(\mathcal{L}(f))(t)$





$F(z) = \mathcal{L}(f(\cdot))(z)$

Inverse stereographic projection

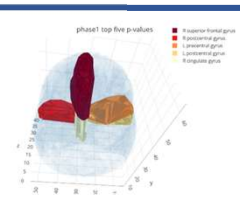


$\text{Reg}(F)(z) = \text{Reg}(\mathcal{L}(f))(z)$

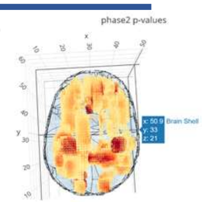
Shen et al., 2024 | Zhang et al., 2022 | Dinov & Velev (2021)

20

Example: Tensor-based Linear Modeling of fMRI



Step 1: ROI analysis



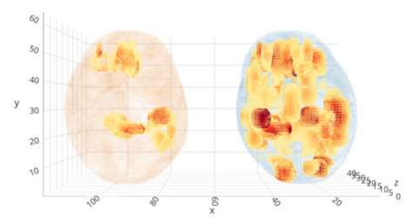
Step 2: Voxel analysis

3-Step Analysis: registering the fMRI data into a brain atlas space, 56 ROIs, tensor linear modeling, post-hoc FDR processing & selection of large clusters of significant voxels are identified within the important ROIs:

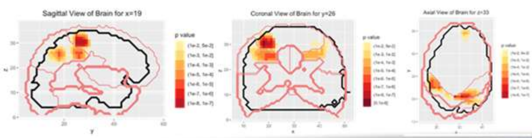
$$Y = \underbrace{\langle X, B\rangle}_{\text{tensor product}} + E$$

The dimensions of the time-tensor Y are $\overbrace{160}^{\text{time}} \times \overbrace{a \times b \times c}^{\text{ROI b}-\text{bo}}$, where the tensor elements represent the response variable $Y[t, x, y, z]$, i.e., fMRI intensity. For fMRI magnitude (real-valued signal), the design kime-tensor X dimensions are:

$$\underbrace{10 * 8}_{\text{Kime}(\text{Time}*e^{i\times Repeat})} \times \underbrace{State}_{\text{Stim vs. Rest (2)}} \times \underbrace{4}_{\text{effects}} \times \underbrace{1}_{\mathbb{R}}.$$



Voxel-based TLM/Analysis
Corrected (step 3, left) vs. Raw (step 2, right)



Step 3: 2D voxel analysis projections (finger-tapping task modeling)

21

Bayesian Inference Representation

- Suppose we have a single spacetime observation $X = \{x_{i_o}\} \sim p(x \mid \gamma)$ and $\gamma \sim p(\gamma \mid \varphi = \text{phase})$ is a process parameter (or vector) that we are trying to estimate.
- Spacekime analytics aims to make appropriate inference about the process X.
- The sampling distribution, $p(x \mid \gamma)$, is the distribution of the observed data X conditional on the parameter γ and the prior distribution of the parameter γ before the observing the data is $p(\gamma \mid \varphi)$, where $\varphi = \text{phase aggregator}$.
- Assume that the hyperparameter (vector) φ, which represents the kime-phase estimates for the process, can be estimated by $\hat{\varphi} = \varphi'$.
- Such estimates may be obtained from an oracle (model distribution), approximated using similar datasets, acquired as phases from samples of analogous processes, derived via some phase-aggregation strategy, or analytically computed (e.g., via Laplace transform).
- Let the posterior distribution of the parameter γ given the observed data $X = \{x_{i_o}\}$ be $p(\gamma|X, \varphi')$ and the process parameter distribution of the kime-phase hyperparameter vector φ be $\gamma \sim p(\gamma \mid \varphi)$.

22

Bayesian Inference Representation

- We can formulate spacekime inference as a Bayesian parameter estimation problem:

$$\underbrace{p(\gamma|X,\varphi')}_{\text{posterior distribution}} = \frac{p(\gamma,X,\varphi')}{p(X,\varphi')} = \frac{p(X|\gamma,\varphi')\times p(\gamma,\varphi')}{p(X,\varphi')} = \frac{p(X|\gamma,\varphi')\times p(\gamma,\varphi')}{p(X|\varphi')\times p(\varphi')} =$$

$$\frac{p(X|\gamma,\varphi')}{p(X|\varphi')}\times\frac{p(\gamma,\varphi')}{p(\varphi')} = \frac{p(X|\gamma,\varphi')\times p(\gamma|\varphi')}{\underbrace{p(X|\varphi')}_{\text{observed evidence}}} \propto \underbrace{p(X|\gamma,\varphi')}_{\text{likelihood}}\times\underbrace{p(\gamma|\varphi')}_{\text{prior}}.$$

- In Bayesian terms, the posterior probability distribution of the unknown parameter γ is proportional to the product of the likelihood and the prior.
- In probability terms, the posterior = likelihood × prior, divided by the observed evidence, in this case, a single spacetime data point, x_{i_o}.

23

Bayesian Inference Representation

- Spacekime analytics based on a single spacetime observation x_{i_o} can be thought of as a type of Bayesian prior-predictive *or* posterior-predictive distribution estimation problem
  - Prior predictive distribution of a new data point x_{j_o}, marginalized over the *prior* – i.e., the sampling distribution $p(x_{j_o}|\gamma)$ weight-averaged by the pure *prior* distribution):

$$p(x_{j_o}|\varphi') = \int p(x_{j_o}|\gamma)\times\underbrace{p(\gamma|\,\varphi')}_{\text{prior distribution}}\,d\gamma\ .$$

  - Posterior predictive distribution of a new data point x_{j_o}, marginalized over the *posterior*; i.e., the sampling distribution $p(x_{j_o}|\gamma)$ weight-averaged by the *posterior* distribution:

$$p(x_{j_o}|x_{i_o},\varphi') = \int p(x_{j_o}|\gamma)\times\underbrace{p(\gamma|x_{i_o},\varphi')}_{\text{posterior distribution}}\,d\gamma\ .$$

- The difference between these two predictive distributions is that
  - The posterior predictive distribution is updated by the observation $X=\{x_{i_o}\}$ and the hyperparameter, φ (phase aggregator),
  - The prior predictive distribution only relies on the values of the hyperparameters that appear in the prior distribution

24

Bayesian Inference Simulation

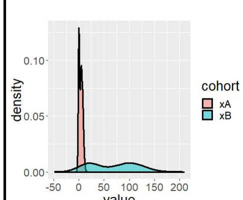
- Simulation example using 2 random samples drawn from mixture distributions each of $n_A = n_B = 10\text{K}$ observations
  1) $\{X_{A,i}\}_{i=1}^{n_A}$, where $X_{A,i} = 0.3U_i + 0.7V_i$, $U_i \sim N(0,1)$ and $V_i \sim N(5,3)$, and
  2) $\{X_{B,i}\}_{i=1}^{n_B}$, where $X_{B,i} = 0.4P_i + 0.6Q_i$, $P_i \sim N(20,20)$ and $Q_i \sim N(100,30)$.

- The intensities of cohorts A and B are independent and follow different mixture distributions. We'll split the first cohort (A) into training (C) and testing (D) subgroups, and then
  1) Transform all four cohorts into Fourier k-space,
  2) Iteratively randomly sample single observations from the (training) cohort C,
  3) Reconstruct the data into spacetime using a single kime-magnitude value and alternative kime-phase estimates derived from cohorts B, C, and D, and
  4) Compute the classical spacetime-derived population characteristics of cohort A and compare them to their spacekime counterparts obtained using a single C kime-magnitude paired with B, C, or D kime-phases.

25

Bayesian Inference Simulation

Summary statistics for the original process (cohort A) and the corresponding values of their counterparts computed using the spacekime reconstructed signals based on kime-phases of cohorts B, C, and D. The estimates for the latter three cohorts correspond to reconstructions using a single spacetime observation (i.e., single kime-magnitude) and alternative kime-phase priors (in this case, kime-phases derived from cohorts B, C, and D).



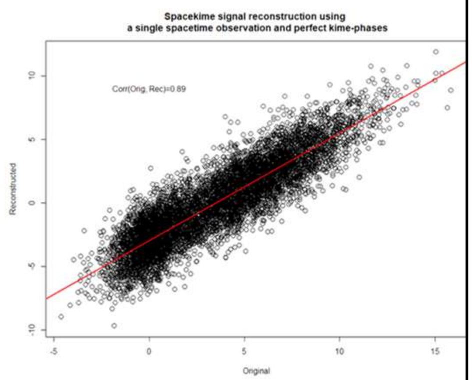
	Spacetime	Spacekime Reconstructions (single kime-magnitude)		
Summaries	**(A)** **Original**	(B) Phase=Diff. Process	$(C$, training) Phase=True	$(D$, testing) Phase=Independent
Min	-2.38798	-3.798440	-2.98116	-2.69808
1st Quartile	-0.89359	-0.636799	-0.76765	-0.76453
Median	0.03311	0.009279	-0.05982	-0.08329
Mean	**0.00000**	**0.000000**	**0.00000**	**0.00000**
3rd Quartile	0.75772	0.645119	0.72795	0.69889
Max	3.61346	3.986702	3.64800	3.22987
Skewness	0.348269	0.001021943	0.2372526	0.31398
Kurtosis	**-0.68176**	**0.2149918**	**-0.4452207**	**-0.3270084**

26

Bayesian Inference Simulation

The correlation between the original data (A) and its reconstruction using a single kime magnitude and the correct kime-phases (C) is $\rho(A, C) = 0.89$.

This strong correlation suggests that a substantial part of the A process energy can be recovered using only a single observation. In this case, to reconstruct the signal back into spacetime and compute the corresponding correlation, we used a single kime-magnitude (sample-size=1) and process C kime-phases.



Bayesian Inference Simulation

Let's demonstrate the Bayesian inference corresponding to this spacekime data analytic problem using a simulated bimodal experiment:

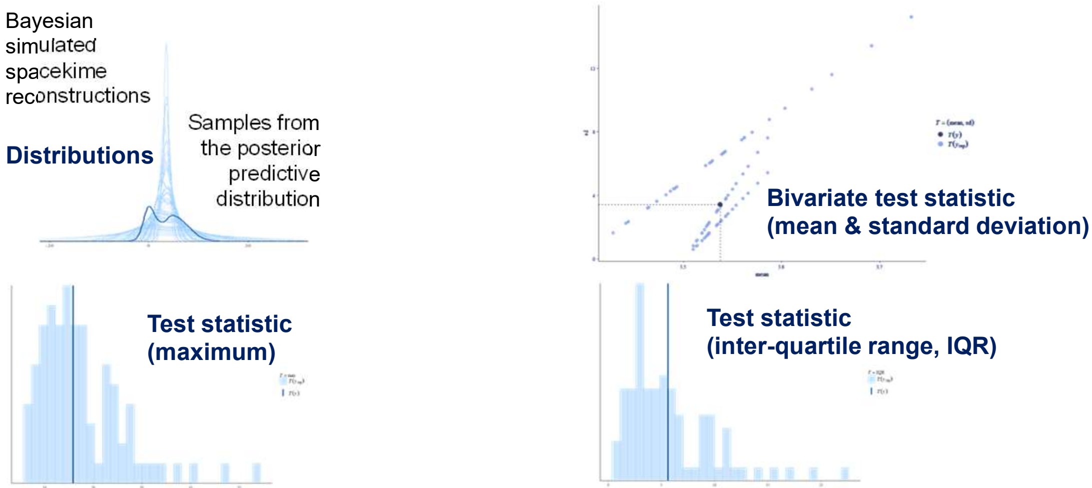
$$X_A = 0.3U + 0.7V, \text{ where } U \sim N(0,1) \text{ and } V \sim N(5,3)$$

Specifically, we will illustrate the Bayesian inference using repeated single spacetime observations from cohort A, $X = \{x_{i_o}\}$, and varying kime-phase priors ($\varphi = \text{phase aggregator}$) obtained from cohorts B, C, or D, using different posterior predictive distributions.

Relations between the empirical data distribution (**dark blue**) and samples from the posterior predictive distribution, representing Bayesian simulated spacekime reconstructions (**light-blue**). The derived Bayesian estimates do not perfectly match the empirical distribution of the simulated data, yet there is clearly information encoding that is captured by the spacekime data reconstructions.

This signal compression can be exploited by subsequent model-based or model-free data analytic strategies for retrospective prediction, prospective forecasting, ML classification, AI derived clustering, and other spacekime inference methods.

Bayesian Inference Simulation

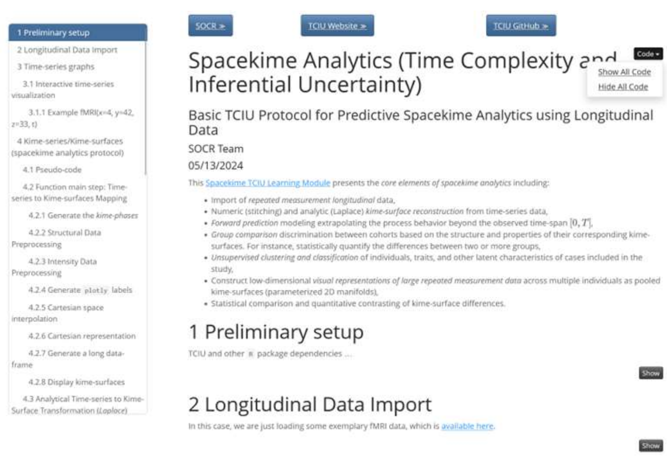


Empirical data distribution (**dark blue**) & samples from the **posterior predictive distribution** Bayesian spacekime reconstructions (**light-blue**).

29

Part 2: Hands-on Spacekime Analytics Tutorial

[TCIU/Spacekime Analytics Tutorial: Basic TCIU Protocol for Predictive Spacekime Analytics using Longitudinal Data]



https://www.socr.umich.edu/TCIU/HTMLs/Chapter6_TCIU_Basic_SpacekimePredictiveAnalytics.html

30

Acknowledgments

Funding

- NIH: UL1 TR002240, R01 CA233487, R01 MH121079, R01 MH126137, T32 GM141746
- NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- **SOCR**: Yueyang Shen, Zerihun Bekele, Milen Velev, Kaiming Cheng, Shihang Li, Daxuan Deng, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang, Yuxin Wang, Rongqian Zhang, Yuyao Liu, Yupeng Zhang, Yunjie Guo, Simeone Marino
- **SPL/HBBS/DCMB/MIDAS/MCAIM** Centers: Dana Tschannen, Chris Anderson, Michelle Aebersold, Maureen Sartor, Josh Welch, Maryam Bagherian, Lydia Bieri, Kayvan Najarian, Chris Monk, Issam El Naqa, Brian Athey