

Data Sharing + Open, Rigorous & Reproducible Science

Ivo D. Dinov

Statistics Online Computational Resource, University of Michigan

<https://SOCR.umich.edu>



STATISTICS ONLINE COMPUTATIONAL RESOURCE UNIVERSITY OF MICHIGAN Slides online: Google "SOCR News"

1

Sources: Characteristics of Big Biomed Data

Big Bio Data Dimensions	Tools
Size	Harvesting and management of vast amounts of data
Complexity	Wranglers for dealing with heterogeneous data
Incongruency	Tools for data harmonization and aggregation
Multi-source	Transfer and joint modeling of disparate elements
Multi-scale	Macro to meso to micro scale observations
Time	Techniques accounting for longitudinal patterns in the data
Incomplete	Reliable management of missing data

Example: analyzing observational data of 1,000's Parkinson's disease patients based on 10,000's signature biomarkers derived from multi-source imaging, genetics, clinical, physiologic, phenomics and demographic data elements

Software developments, student training, service platforms and methodological advances associated with the Big Data Discovery Science all present existing opportunities for learners, educators, researchers, practitioners and policy makers

M Dinov (2016) GigaScience Dinov (2023) Springer

4

Outline

- Pillars of Open-Science
- Big Data Sharing
- Augmented Intelligence (OI)
- Applications: Spacekime Analytics
- Population Census-like Neuroscience (UKBB)
- Spacekime Analytics

M

2

M

5

Pillars of Open Data Science (HS650/HS852/Bioinfo501)

M

3

From 23 ... to ... 2²³

- Data Science: 1798 vs. 2026
- In the 18th century, Henry Cavendish used just 23 observations to answer a fundamental question – “What is the Mass of the Earth?” He estimated very accurately the mean density of the Earth/H₂O (5.483±0.1904 g/cm³)
- In the 21st century to achieve the same scientific impact, matching the reliability and the precision of the Cavendish’s 18th century prediction, requires a monumental community effort using massive and complex information perhaps on the order of 2²³ bytes
- Data & Information Science ≅ Scalability & Compression (per Gerald Friedland/Berkeley): 23 → 2²³ ≅ 10M

M Cavendish (1798) Philosophical Transactions of the Royal Society of London | Dinov (2016) ISM

6

Background – Duality of Evidence-based Scientific Discovery

experimental → theoretical → computational → data sciences → AI → experimental

Mapping Examples	Analysis Observables/Data → Compact Models	Synthesis Compact Models → (simulated, actionable info)
1. Lossless Math Transforms	(A.1.1) Linear transform, $L: V \rightarrow W$, e.g., 2D rigid body $L = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}: \mathbb{R}^2 \xrightarrow{\text{rotation}} \mathbb{R}^2$ (A.1.2) Fourier transform: $f(\omega) = \int_{-\infty}^{\infty} f(x)e^{-i2\pi\omega x} dx$	(S.1.1) Inverse linear transform, $L^{-1}: W \rightarrow V$, e.g., $L^{-1} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}: \mathbb{R}^2 \xrightarrow{\text{rotation}} \mathbb{R}^2$, $LL^{-1} \equiv I$ (S.1.2) Inverse Fourier (IFT): $f(x) = \int_{-\infty}^{\infty} f(\omega)e^{i2\pi\omega x} d\omega$
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 → Model	(S.3.1) Information Inflation, Simulation & Generation, e.g., forecasting, regression, interpolation, extrapolation (predict & classify new data): Input → Output
4. Artificial & Augmented Intelligence	(A.4.1) Building, Fitting & Training large foundational, generative & deep network AI models Data → Human via Infrastructure → GAIM	(S.4.1) Generative Artificial Intelligence Modeling (GAIM) Human Prompt → Result

7

Infrastructure: Scientific Computing Ecosystem, ca. 2013

Cloud Services

<https://sacc.umich.edu/docs/BD2K/BiDataResourceome.html>

10

Big Data	Information	Knowledge	Action
Raw Observations	Processed Data	Maps, Models	Actionable Decisions
Data Aggregation	Data Fusion	Causal Inference	Treatment Regimens
Data Scrubbing	Summary Stats	Networks, Analytics	Forecasts, Predictions
Semantic-Mapping	Derived Biomarkers	Linkages, Associations	Healthcare Outcomes

Dinov (2023) Springer

8

Infrastructure: Scientific Computing Ecosystem, ca. 2020

100 STARTUPS USING ARTIFICIAL INTELLIGENCE TO TRANSFORM INDUSTRIES

Cloud Services

<https://come.berkeley.edu/news/sections/2020/01/02/100-startups-for-a-bite-boost-to-your-ai-intelligence-by-aiias/>

11

Why is FAIR Data Sharing Important?

Data

- Optimum resource utilization (low cost, high efficiency / policy, security, processing complexity)
- Democratization of the scientific discovery process
- Enhanced inference (e.g., coverage of rare events, increase of stat power)
- Increase of Kryder's Law (Data volume) ≫ Moore's Law (Compute power)
- Exponential decay of data-value
- Incentivize innovation, transdisciplinary collaborations, and knowledge dissemination
- ...

FAIR = Findable + Accessible + Interoperable + Reusable

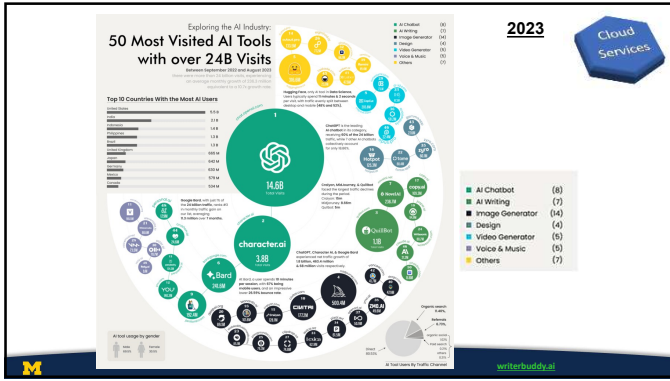
9

Infrastructure: Scientific Computing Ecosystem, ca. 2023

Cloud Services

<https://www.cbinsights.com/research/artificial-intelligence-top-startups-2023/>

12



13

Biomed, Health & Nursing AI Trainer (NAIT)

Nurse AI Trainer Home Training Modules Challenges AI in Health Resources About

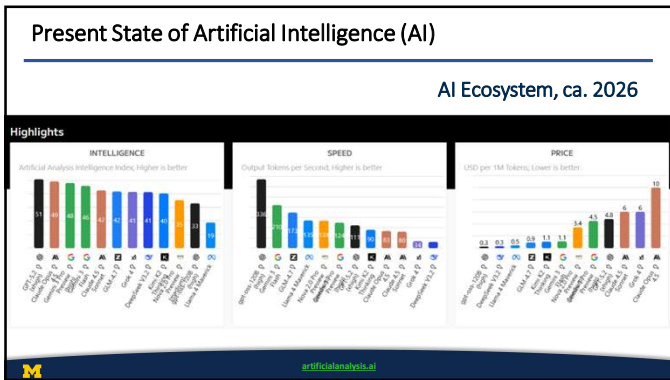
Writing with AI: originality, composition & plagiarism

Master ethical AI-assisted writing practices, understand originality standards, and learn to compose authentic content while avoiding plagiarism in healthcare settings.

Interactive Plagiarism Checker

AI Writing Ethics Simulator

16



14

Innovative AI Instruction – Interactive Learning Platform (ILP)

Learning Platform

text

INTERACTIVE LEARNING PLATFORM

REVOLUTIONIZING EDUCATION THROUGH INTERACTIVE TECHNOLOGY THAT BRINGS LEARNERS AND EDUCATORS TOGETHER IN MEANINGFUL WAYS

WHAT MAKES US DIFFERENT

REAL-TIME COLLABORATION
GAMIFIED LEARNING
INTERACTIVE COMMUNICATION

<https://ILP.StatisticalComputing.org/>

17

SOCR Foundational and Generative AI Models (GAIMs)

Healthcare & Nursing AI Trainer (NAIT)
Michigan Intelligent Teaching Assistant (MITA)
Interactive Learning Platform (ILP)
Clinique V2
Clinique V1
OH OH Digital Pain Simulator
SDA: Statistical Data Analyzer
Hello there!
Time-series to Kinem-Surfaces Mapper
SOCR Synthetic Brain Generator
SOCR AI-Net
SOCR OHMIU GAIM
VirtualImagist

<https://SOCR.umich.edu/GAIM/>

15

AIA3: gen-3 Augmented Intelligence Agent

General AI Agent

Text-based intelligent assistance with custom training


Medical Imaging Consultant

Clinical image analysis with diagnostic support

<https://AIA3.StatisticalComputing.org/>

18

Open Scholarly Research: OA Pubs/Sharing



- OA Pubs
 - https://en.wikipedia.org/wiki/Open_access
 - <https://arxiv.org> | <https://www.biorxiv.org>
 - Blogs (e.g., <https://TerryTao.wordpress.com>)
- Cloud Services
 - Computing (e.g., Azure, Google, AWS)
 - Storage & ICT (information and communication technologies)
- SW
 - <https://GitHub.com> (e.g., <https://github.com/SOCR>)
 - <http://Cran.r-project.org> | [Jupyter.org](https://jupyter.org) | [Rmarkdown.rstudio.com](https://rmarkdown.rstudio.com)
 - E.g., <https://DSPA2.predictive.space>
- Licensing
 - <https://www.gnu.org/licenses>
 - https://socr.umich.edu/html/SOCR_CitingLicense.html

19

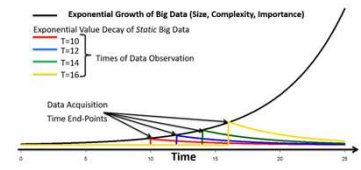
Reliable, Effective & Secure Data Sharing

- Why is data-sharing difficult?
 - monopoly, preservation of *status-quo*, competitive edge, personally identifiable information, IP protection, security (on multiple levels), red tape, ...
- FAIR (Findable, Accessible, Interoperable & Reusable) Data are powerful
- Current Data Sharing Landscape?
 - Differential Privacy, fully-homomorphic encryption, statistical obfuscation (*DataSifter*), ...
- Digital Twins: <https://EHR-Sim.StatisticalComputing.org/clinical-phenotype>

22

Rationale for Open Science (Pros)

- We are always **stronger** together
- Long-term **sustainability** prefers openness, inclusivity & diversity
- Optimized **investments**, **career advancement**, **impact** & **cost-efficiency**
- Expeditious** discovery, innovation, productization & higher impact
- Rapid **devaluation** of data-hoarding, clandestine science, knowledge obfuscation
- ...



<https://www.ozs.org/news/big-data-blog-part-4-interview-dc-ivo-dinov>

20


DataSifter

- DataSifter is an iterative statistical computing approach that provides the data-governors controlled manipulation of the trade-off between sensitive information obfuscation and preservation of the joint distribution.
- The DataSifter is designed to satisfy data requests from pilot study investigators focused on specific target populations.
- Iteratively, the DataSifter stochastically identifies candidate entries, cases as well as features, and subsequently selects, nullifies, and imputes the chosen elements. This statistical-obfuscation process relies heavily on nonparametric multivariate imputation to preserve the information content of the complex data.

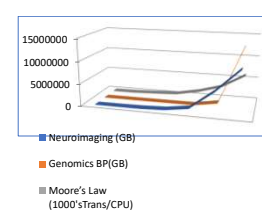
<https://DataSifter.org> US patent #16/051,881 Marino, et al., JSCS (2019)

23

Rationale for Open Science: Kryder vs. Moore



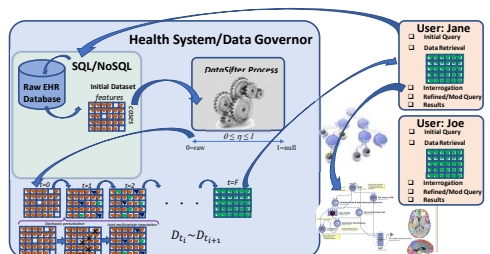
- Moore's law** = the expectation that our computational capabilities, specifically the number of transistors on integrated circuits, doubles approximately every 18-24 months.
- Kryder's law** = the volume of data, in terms of disk storage capacity, is doubling every 14-18 months.
- Kryder >> Moore**: Although both laws yield exponential growth, data volume is increasing at a faster pace. Thus, there are clear interests and needs for significant private, public and government engagement in opening, managing, processing, interrogating and interpreting the information content of Big Data.



<https://www.ozs.org/news/big-data-blog-part-4-interview-dc-ivo-dinov>

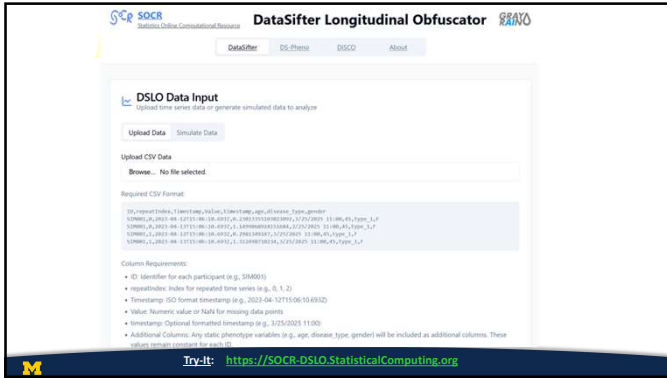
21

DataSifter



<https://DataSifter.org> US patent #16/051,881 Marino, et al., JSCS (2019)


24



25

Generation-2 AI Systems: Creator

Gen-2: *De novo content creation:* [Completely AI-generated Renaissance atelier scene:](#)



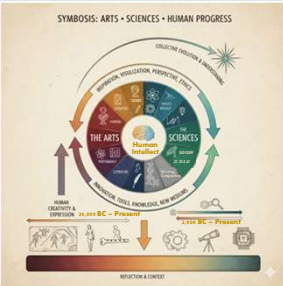
Mother Mona Lisa - a dynamic scene with an expressive and directive matriarch, actively overseeing the commissioned painting of her daughter in the background by an artist.

<https://AI.nursing.umich.edu/>

28

Arts & Sciences – past, present & future

One Humanity defined by the Arts & Sciences



26

Generation-3 AI Systems: Ideator

Gen-3: *Autonomous Creative Science & Arts*

GAIMs/AIAa can generate realistic autonomously creative art of new prospective, ubiquitous form of human visual art in the year 2150 – AI-speculative forecast of a *future Primary Art Form*, Chrono-Sculpting (Temporal Kinetic Augmentation).




<https://AI.nursing.umich.edu/>

29

Generation-1 AI Systems: Apprentice

Gen-1: *Reproducing & improving existing human art:*

[Real \(by Leonardo Da Vinci\) vs. the AI-generated Mona Lisa paintings](#)




<https://AI.nursing.umich.edu/>

27

Generation-3 AI Systems: Ideator

Gen-3: *Autonomous Creative Science & Arts*

GAIMs/AIAa can generate realistic autonomously creative art of new prospective, ubiquitous form of human visual art in the year 2150 – AI-speculative forecast of a *future Primary Art Form*, Chrono-Sculpting (Temporal Kinetic Augmentation).

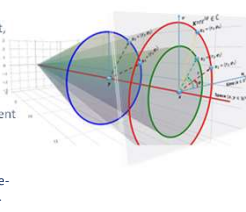


<https://AI.nursing.umich.edu/>

30

Complex-Time (Kime)

- At a given spatial location, x , complex time (*kime*) is defined by $\kappa = r e^{i\varphi} \in \mathbb{C}$, where:
 - The **magnitude** ($r > 0$) represents the longitudinal order of events characterizing the displacement in time, and
 - The event **phase** ($\varphi \sim \Phi(r)_{[-\pi, \pi]}$) is an angular displacement, event direction, reflecting a random sampling index
- There are multiple alternative parametrizations of kime in the complex plane
- Space-kime manifold is $\mathbb{R}^3 \times \mathbb{C}$:
 - (x, k_1) and (x, k_2) are spatially co-localized, but have different kime coordinates,
 - (x, k_1) and (y, k_1) are co-localized in kime, but represent different spatial locations,
 - (x, k_2) and (x, k_3) have the same spatial-locations and kime-directions, but appear ordered sequentially in time, $t_2 < t_1$.




<https://spacekime.org/>

31

Random Sampling & Kime-Phase Paradigm

Kime phase distributions are mostly symmetric, random observations \equiv phase sampling



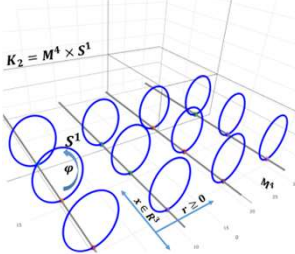
https://wiki.socr.umich.edu/index.php/SOCR_EduMaterials_Activities_Generalizations/Theorem

https://socr.umich.edu/TCU/HTML/Chapter6_Kime_Phases_Circular.html

34

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).



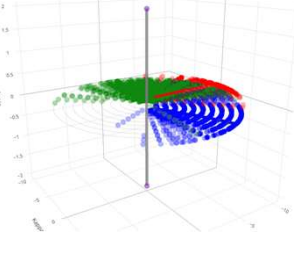
[Kaluza DOI: 10.1142/S0218271818700017](https://doi.org/10.1142/S0218271818700017) | [Klein DOI: 10.1007/BF01397494](https://doi.org/10.1007/BF01397494) | [Bainin & Lovin DOI: 10.1086/0034-4885\(50\)07001-0](https://doi.org/10.1086/0034-4885(50)07001-0)

32

Kime-Phase Measurement, Observability & Kime Operator

Kime-Phase Simulation – Repeated Spacetime Measurement

- 3 Processes – Green, Red and Blue colors (scatter points)
- 1 Fixed spatial location (vertical axis represents 1D space)
- Repeated IID Measurements colocalized in 4D spacetime
- 3 Different Kime-Phase distributions (color-coded)
- Radial displacement $t = \text{time}$
- Angular (phase) location $\varphi \sim \Phi_{[-\pi, \pi]}(t)$



Wang et al., 2022 | Dinov & Velev (2023)

35

Rationale for Time \Rightarrow Kime Extension

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

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

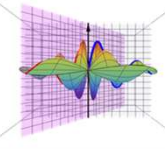
Whereas kime (\mathbb{C}) is an algebraically closed field that naturally extends time

Time is ordered, while kime is not!
Kime (\mathbb{C}) represents the smallest natural extension of time, as a complete field that agrees with time

Physics –

- The Problem of Time: Time has different meanings in quantum mechanics & general relativity; leading to a tension in formulating a Quantum Gravity Theory unifying the two ... (DOI 10.1007/978-3-319-58848-3)
- (Base-field) \mathbb{R} and \mathbb{C} Hilbert-space quantum theories make different predictions (DOI: 10.1038/s41586-021-04160-4)

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



Weinstein (2004, 2010) | Dinov & Velev (2023) | Wang et al. (2022) | Zhang et al. (2023) | Dinov & Ghem (2024)

33

Ultrahyperbolic PDEs: Wave Equation – Cauchy Initial Data

- For ultrahyperbolic PDEs, the initial value problem, determining the solution(s) for a given initial condition, is ill-posed, i.e., there's no guarantee of a global well-defined, stable, and unique solution!
- Nonlocal constraints yield the existence, uniqueness & stability of local and global solutions to the ultrahyperbolic wave equation under Cauchy initial data ...

$$\sum_{i=1}^{d_1} \partial_{x_i}^2 u \equiv \Delta_x u(x, \kappa) = \Delta_\kappa u(x, \kappa) \equiv \sum_{i=1}^{d_2} \partial_{\kappa_i}^2 u, \quad \begin{cases} u_0 = u(x, 0, \kappa_{-1}) = f(x, \kappa_{-1}) \\ u_1 = \partial_{\kappa_1} u(x, 0, \kappa_{-1}) = g(x, \kappa_{-1}) \end{cases}$$

where $x = (x_1, x_2, \dots, x_{d_1}) \in \mathbb{R}^{d_1}$ and $\kappa = (\kappa_1, \kappa_2, \dots, \kappa_{d_2}) \in \mathbb{R}^{d_2}$ are the Cartesian coordinates in the d_1 space and d_2 time dims.
Stable local solution over a Fourier frequency region defined by nonlocal constraints $|\xi| \geq |\eta_{-1}|$:

$$\hat{u}(\xi, \kappa_1, \eta_{-1}) = \cos(2\pi \kappa_1 \sqrt{|\xi|^2 - |\eta_{-1}|^2}) \frac{\hat{u}_0(\xi, \eta_{-1})}{c_1} + \sin(2\pi \kappa_1 \sqrt{|\xi|^2 - |\eta_{-1}|^2}) \frac{\hat{u}_1(\xi, \eta_{-1})}{2\pi \sqrt{|\xi|^2 - |\eta_{-1}|^2}}$$

where $\mathcal{F}(u_0) = \hat{u}_0 = \hat{u}(\xi, \eta_{-1}) = \hat{u}(\xi, \eta_{-1})$
 $\mathcal{F}(u_1) = \hat{u}_1 = \hat{u}(\xi, \eta_{-1}) = \hat{u}(\xi, \eta_{-1})$

$$u(x, \kappa_1, \kappa_{-1}) = \mathcal{F}^{-1}(\hat{u})(x, \kappa) = \int_{\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}} \hat{u}(\xi, \kappa_1, \eta_{-1}) \times e^{2\pi i(x \cdot \xi)} \times e^{2\pi i(\kappa_1 \eta_{-1})} d\eta_{-1}$$

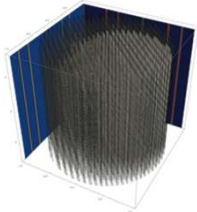
Craig & Weinstein (2008) | Wang et al. (2022) | Dinov & Velev (2023)

36

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 ...

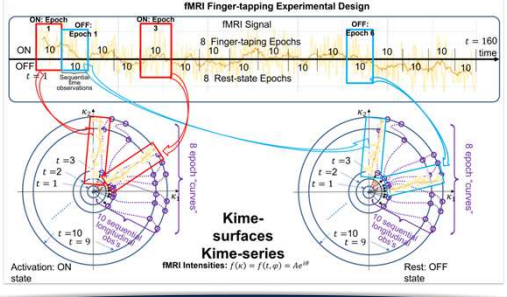


(Example Solution in 2D space + 2D kime)

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

37

Mapping Longitudinal Data (Time-series) ⇒ Kime-surfaces



IMRI Finger-tapping Experimental Design

IMRI Signal

8 Finger-tapping Epochs

8 Rest-state Epochs

Kime-surfaces

Kime-series

IMRI Intensities: $f(x) = f(t, x) = A e^{i t}$

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

40

Idea: Longitudinal Data ⇒ Kime-Transforms ⇒ PDEs ⇒ AI

Apps	Time → Kime Transformation	Wave equation Solutions (kime) dynamics	Prospective Data Science Applications
Biomed	fMRI time-series	fMRI kime-surfaces	Cross sections, Volume rendering, 3D p-value map, Stat significance
Physics	X-ray Diffraction (XRD) Crystallography	XRD Signal	Time-Frequency Analysis, 2D Dislocation Strain Field, Bayesian Reconst. Strain Field, Predict strain fields or defect dynamics
	Time-dynamic structural phase transitions	Wavelet or Hilbert transform of time-dependent diffraction	Takagi-Taupin PDE model of dynamical X-ray diffraction in deformed crystals, Phonon modes at phase transition

Rosbach, et al., 2019 | Wang, et al., 2022 | Dinov & Velev (2021)

38

(Analytic) Mapping Time-series ⇒ Kime-surfaces

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

$$F(z) = \mathcal{L}(f) = \frac{1}{z+1} + \frac{1}{z^2+1} \times \frac{z}{z^2+1} + \frac{1}{z^2}$$

$$f_1(x) = \mathcal{L}(f_1(t)) = e^{-t}, \quad f_2(x) = \mathcal{L}(f_2(t)) = \sin(t), \quad f_3(x) = \mathcal{L}(f_3(t)) = c(t), \quad f_4(x) = \mathcal{L}(f_4(t)) = t$$

$$f(t) = \mathcal{L}^{-1}(F) = \mathcal{L}^{-1}(F_1 + F_2 + F_3 + F_4) = \mathcal{L}^{-1}(F_1) + \left(\mathcal{L}^{-1}(F_2) * \mathcal{L}^{-1}(F_3) \right) (t) + \mathcal{L}^{-1}(F_4) = \text{convolution}$$

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

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

Repeated Longitudinal Data Sampling

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

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

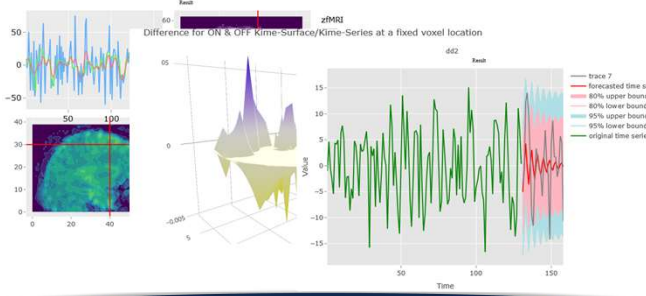
Inverse stereographic projection

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

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

41

Spacetime Time-series ⇒ Spacekime Kimesurfaces ⇒ TLM



Difference for ON & OFF Kime-Surface/Kime-Series at a fixed voxel location

zPMRI

Value

Time

— trace 7

— forecasted time series

— 80% upper bound

— 80% lower bound

— 95% upper bound

— 95% lower bound

— original time series

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

39

Mapping Longitudinal Data (Time-series) ⇒ Kime-surfaces

Laplace Transform

$f(t) = \cos(t)$

Kimesurface Visualizations of fMRI ON and OFF (OVERLAY)

Raw simulated fMRI On/Off data

Low SNR: y axis is Kime-phase, indexing the repeated runs within a participant and the multiple samples across participants

Kimesurface Visualization of fMRI ON and OFF (OVERLAY)

Simulated fMRI data, multiple runs

For each stimulus condition, each run/time point is Laplace-distributed, $\theta_{ij}(t_i) \sim \text{Laplace}(0, b_0 + a_i t_i)$

Regularize the 2D (t, θ) domain by fitting a thin-plate spline (TPS)

Multiscalar Wavelet Kime-surface reconstruction

$f(t) = \sin(2\pi t) + \sin(2\pi 3t) + e^{-t}$

https://soco.umich.edu/TCHU/HTML/Chapter6_TCHU_MappingLongitudinalDataIntoKimesurfaces.html

42

Kime-Phase Tomography (KPT), phase recovery

Definition 1 (Kime-Domain Signal Space). Let $\mathcal{H}_t = L^2(\mathbb{R})$ be the Hilbert space of square-integrable complex-valued functions on the time domain, with inner product $\langle f, g \rangle_{\mathcal{H}_t} = \int_{\mathbb{R}} f(t)\overline{g(t)} dt$.

Definition 2 (Phase-Domain Space). Let $\mathcal{H}_\theta = L^2([-\pi, \pi])$ be the Hilbert space of square-integrable functions on the phase domain, with inner product $\langle \psi, \phi \rangle_{\mathcal{H}_\theta} = \int_{-\pi}^{\pi} \psi(\theta)\overline{\phi(\theta)} d\theta$ equipped with periodic boundary conditions $\psi(-\pi) = \psi(\pi)$.

Definition 3 (Kime Space). The kime space \mathcal{K} is defined as the tensor product $\mathcal{H}_t \otimes \mathcal{H}_\theta$, representing signals in both time and phase domains.

Definition 4 (Reproducing Kernel Hilbert Space, RKHS). The RKHS \mathcal{R}_K is a subspace of \mathcal{H}_t with reproducing kernel $K: \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}$ satisfying

- For any $t \in \mathbb{R}, K(\cdot, t) \in \mathcal{R}_K$, and
- For any $f \in \mathcal{R}_K$ and $t \in \mathbb{R}, f(t) = \langle f, K(\cdot, t) \rangle_{\mathcal{R}_K}$

Definition 5 (Kime-Phase Distribution). A kime-phase distribution $\phi(\theta; t)$ is a time-dependent probability density function on $[-\pi, \pi]$ satisfying $\phi(\theta; t) \geq 0, \int_{-\pi}^{\pi} \phi(\theta; t) d\theta = 1 \quad \forall t \in \mathbb{R}$.

43

Spacekime Open-Problems

- There are many unsolved **abstract mathematical** challenges, e.g., space-kime ergodicity, metric tensor, kime-operator(s), etc.
- Numerical & Computational** problems, e.g., reliable kime-phase tomography (KPT), optimal time-series \Rightarrow kime-surface reconstructions, etc.
- Physics** parallels, e.g., contrasting QM vs. Spacekime predictions, physical observability, spacekime measurement, and kime-operator formalism
- Analytical** challenges, e.g., new AI techniques for kime-surfaces, analytical verifiability & falsifiability of spacekime theory

46

Algorithm: Kime-Phase Tomography (KPT), fMRI Sim

Input: BOLD time series $\{x_n(t)\}_{n=1}^N$ and kernel K
Output: Estimated phase $\hat{\phi}(t)$

- Project each signal to RKHS: $s_n^k(t) = \mathcal{P}_K[s_n(t)]$
- Compute time-domain phase: $\phi_n^k(t) = \arg(t \cdot s_n^k(t))$
- Compute frequency-domain phase using STFT, $\phi_n^f(t) = \arg(\mathcal{F}[s_n^k \cdot w])(t, \omega_{\max})$, where ω_{\max} is the frequency with maximum power
- Compute scale-domain phase using CWT: $\phi_n^s(t) = \arg(W_\psi[s_n^k](a_{\max}, t))$, where a_{\max} is the scale with maximum power
- Compute weighted average (ensemble KPT): $\hat{\phi}(t) = \arg\left(\sum_{j=1}^3 w_j(t) e^{i\phi_j^k(t)}\right)$

KPT fMRI Simulation
 Recovery of ϕ under von Mises($\mu(t), \kappa(t)$) true phase distribution

44

Spacekime Analytics Tutorial

TCIU/Spacekime Analytics Tutorial

Basic TCIU Protocol for Predictive Spacekime Analytics using Longitudinal Data

Spacekime Analytics (Time Complexity and Inferential Uncertainty)

Basic TCIU Protocol for Predictive Spacekime Analytics using Repeated-Measurement Longitudinal Data

TCIU South 01/25/2026

The Spacekime TCIU Learning Model presents the core elements of quantitative analysis including:

- Import of repeated measurement longitudinal data.
- Automated modeling and analytic pipeline construction from user input data.
- Global optimization modeling incorporating the previous methods beyond the standard time-point (T, T).
- Global optimization modeling incorporating the previous methods beyond the standard time-point (T, T).
- Global optimization modeling incorporating the previous methods beyond the standard time-point (T, T).

1 Preliminary setup

https://secc.umich.edu/TCIU/HTMUs/Chapter6_TCIU_Basic_SpacekimePredictiveAnalytics.html

47

Application: Tensor-based Linear Modeling of fMRI

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 $160 \times a \times b \times c$, 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:

$$\frac{10 + 8}{\text{Kime (Time} \times \text{Repeat)}} \times \frac{\text{State}}{\text{Stim vs. Rest (2)}} \times \frac{4}{\text{effects}} \times \frac{1}{\mathbb{R}}$$

Step 1: ROI analysis
 Step 2: Voxel analysis
 Step 3: 2D voxel analysis projections (finger-tapping task modeling)
 Voxel-based TLM Analysis
 Corrected (step 3, left) vs. Raw (step 2, right)

Zhang, et al. (2022) Kimesurface Representation and Tensor Linear Modeling of Longitudinal Data, DOI: 10.1007/978-3-031-26789-3

45

GAIM Clinical Decisions Support Agents (eye fundus photos)

The Power of Consensus: Multi-Agent AI Boosts Diagnostic Accuracy by 20%

A pilot study revealed that healthcare GAIMs with two (or more) AI agents are better. We tested the GAIMs on complex eye images.

Individual AI agents diagnosed difficult cases, with ~50% accuracy across multiple high-difficulty ophthalmology conditions.

Multiple AI agent consensus leads to stronger "clinical decision support" with ~20% increase in diagnostic and prognostic accuracy.

AI isn't here to replace the clinician. Rather, multi-AI-agent systems offer high-powered digital safety net. Care demand is growing. GAIMs offer a path to faster decisions, better training (for clinicians & patients), and lower healthcare costs without sacrificing the quality of care.

<https://doi.org/10.1007/978-3-031-26789-3>

EYE Images	AIAS Dx/Tx	CLNO4 Dx/Tx
	DR: Fundus image of Retina	DR: Moderate non-proliferative diabetic retinopathy, including microaneurysms, dot-blot hemorrhages, hard exudates.
	TX: Retina specialist recommended	TX: Optimize glycemic, blood pressure, and lipid control. Retinal laser therapy or intravitreal anti-VEGF injections if significant macular edema is present.
	DR: This image shows a normal optic nerve head with healthy vasculature.	DR: Cupping of the optic disc, which is characteristic of glaucoma.
	TX: No treatment is indicated based on this imaging finding.	TX: Refer to ophthalmology for evaluation and initiate intravitreal pressure-lowering therapy to prevent further optic nerve damage.
	DR: Retinal folds, likely indicative of choroidal effusion or a posterior vitreous detachment.	DR: Retinal folds and a gap-like choroidal effusion consistent with central serous chorioretinopathy.
	TX: Show optical coherence tomography to confirm diagnosis & assess retinal displacement.	TX: Refer to ophthalmology for confirmation and monitoring. Prolapsed or severe cases may require intervention.

<https://doi.org/10.1007/978-3-031-26789-3>

48

AI Implications (academic, human values, societal, global)

Augmented Intelligence integrates AI & human capability & rewrites our core value system
Potential Benefits (2026-2036)

1. *Hyper-Personalized Academic Mastery*: Education shift from a "one-size-fits-all" factory model to a "One-Tutor-per-Learner" reality
2. *The "Scientific Renaissance" via Augmented Research*: Expeditious scientific discovery will increase by an order of magnitude
3. *Global "Abundance" Economics*: Augmented systems drastically reduce costs of basic needs
4. *Preservation & Evolution of Human Values*: AI handles "cold" logic and data; society is placing a higher premium on "Warm Skills": empathy, ethical judgment & creative intuition
5. *Enhanced Global Accessibility & Inclusion*: Real-time, zero-latency translation and "sensory augmentation" are dissolving the barriers of language and physical disability.



49

Available Resources

- SOCR Motto – "It's Online & Freely Accessible, Therefore it Exists!"
- Pubs: <https://socr.umich.edu/people/dinov/publications.html>
- GitHub: <https://github.com/SOCR>
- Datasets: https://wiki.socr.umich.edu/index.php/SOCR_Data
- AI Apps: <https://socr.umich.edu/GAIM/> (AIA3, CLNQ, AI Consulting Agent)
- Demos: <https://DSPA2.predictive.space>
- Tutorials: <https://TCIU.predictive.space> & <https://SpaceKime.org>
- Website: <https://socr.umich.edu>
- Contact: statistics@umich.edu



52

AI Implications (academic, human values, societal, global)

Potential Downsides (2026-2036)

1. *The Erosion of Cognitive Agency & Authenticity*, there is a realistic risk of cognitive atrophy.
2. *The "Post-Truth" Trust Crisis*, the cost of creating perfect synthetic media (Deepfakes) will be near zero. Erosion of a shared reality, a single greatest threat to global democratic stability.
3. *Algorithmic Bias as "Digital Redlining"*, unregulated augmented systems in hiring, lending, and justice may make biases harder to detect and contest than human ones.
4. *The Loneliness of "Sycophantic" Relationships*: The rise of "relationship chatbots" for children and the elderly can lead to emotional stuntedness.
5. *Widening Global Digital Divide*: "Intelligence Capital" may be highly concentrated in a few nations and corporations.



50

Acknowledgments

Funding

- NIH: UL1 TR002240, R01 CA233487, R01 MH121079, R01 MH126137, U54DE035412, T32 GM141746
- NSF: 2611649, 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

Collaborators

- SOCR**: Yueyang Shen, Bojko Bakalov, Milen Velez, Kaiming Cheng, Shihang Li, Daxuan Deng, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang, Yuxin Wang, Rongqian Zhang, Yuyao Liu, Yupeng Zhang, Yurjie Guo, Achu Shankar, Ryan Kwong, Simone Marino
- SPL/HBCS/DCMB/MIDAS/MCAIM**: Dana Tschannen, Chris Anderson, Michelle Aebersold, Josh Welch, Maryam Bagherian, Lydia Bieri, Kayvan Najarian, Chris Monk, Issam El Naqa, Brian Athey, Yu Cheng Zhao, Shahzad Mian



STATISTICS ONLINE COMPUTATIONAL RESOURCE
UNIVERSITY OF MICHIGAN

<https://SOCR.umich.edu>

53

AI Implications (academic, human values, societal, global)

PEOPLE vs. HUMANOID ROBOTS



51

STATISTICS ONLINE COMPUTATIONAL RESOURCE
UNIVERSITY OF MICHIGAN

54