# Quantum Physics Interface to Data Science, Artificial Intelligence & Spacekime Analytics

Ivo D. Dinov

**Statistics Online Computational Resource**
Health Behavior & Biological Sciences
Computational Medicine & Bioinformatics
Michigan Institute for Data Science
University of Michigan

https://SOCR.umich.edu

Joint work with Milen V. Velev (BTU) & Yueyang Shen (UM)

Based on the book *"Data Science: Time Complexity, Inferential Uncertainty & Spacekime Analytics"*

STATISTICS ONLINE COMPUTATIONAL RESOURCE (SOCR)
UNIVERSITY OF MICHIGAN

Slides Online: "SOCR News"

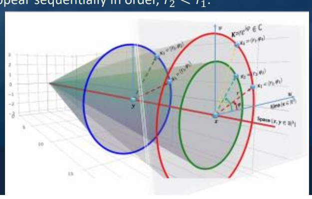
# Outline

- ❑ Complex-Time (*kime*) & Rationale
- ❑ Solutions of untrahyperbolic wave equations
- ❑ Open Spacekime Problems

- ❑ Data/Neuro Science Applications
  - ❑ Random Sampling vs. Hidden Variables Paradigm
  - ❑ Neuroimaging (fMRI): time-series ➔ kime-surfaces
- ❑ Bayesian Formulation of Spacekime Inference
- ❑ Live Demo Links

# 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, or event direction
- ❑ 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, $r_2 < r_1$.



# Rationale for *Time* → *Kime* Extension

- ❑ **Math** – $Time$ is a special case of *kime*, $\kappa = |\kappa|e^{i\varphi}$ where $\varphi = 0$ (nil-phase)
  - algebraically a *multiplicative* (algebraic) group, (multiplicative) unity (identity) $= 1$
  - multiplicative inverses, multiplicative identity, associativity $t_1 * (t_2 * t_3) = (t_1 * t_2) * t_3$
  - The $time$ domain ($\mathbb{R}^+$) is **not** a complete *algebraic field* $(+,*)$:
    - Additive unity (0), element additive inverse $(-t)$: $t + (-t) = 0$; is outside $\mathbb{R}^+$ (time-domain)
    - $x^2 + 1 = 0$ has no solutions in time (or in $\mathbb{R}$) ….

$$\text{Group}(*) \subseteq Ring\left(\underbrace{\overbrace{(+,*)}^{\text{Compatible operations}}}_{\text{associative \& distributive}}\right) \subseteq Field\ \overbrace{(+,*)}^{\text{Group}(+)}$$

  - Classical time ($\mathbb{R}^+$) is a *positive cone* over the field of the real numbers ($\mathbb{R}$)
  - Time forms a subgroup of the multiplicative group of the reals
  - Whereas kime ($\mathbb{C}$) is an algebraically *closed prime field* that naturally extends time
  - *Time* is ordered & *kime* is not – the kime magnitude preserves the intrinsic time order
  - Kime ($\mathbb{C}$) represents the smallest natural extension of time, complete filed that agrees with time
  - The $time$ group is closed under addition, multiplication, and division (but not subtraction). It has the topology of $\mathbb{R}$ and the structure of a multiplicative topological group ≡ additive topological semigroup
- ❑ **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)
- ❑ **AI/Data Science** – Random IID sampling, Bayesian reps, tensor modeling of $\mathbb{C}$ kimesurfaces, novel analytics

Dinov & Velev (2021)

# 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 …

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

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 $|\boldsymbol{\xi}| \ge |\boldsymbol{\eta}_{-1}|$ :

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

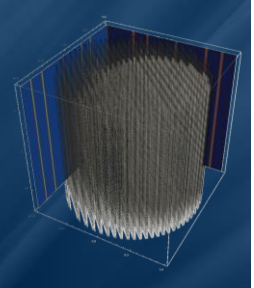
where $\mathcal{F}\begin{pmatrix} u_o \\ u_1 \end{pmatrix} = \begin{pmatrix} \hat{u}_o \\ \hat{u}_1 \end{pmatrix} = \begin{pmatrix} \hat{u}_o(\boldsymbol{\xi}, \boldsymbol{\eta}_{-1}) \\ \hat{u}_1(\boldsymbol{\xi}, \boldsymbol{\eta}_{-1}) \end{pmatrix} = \begin{pmatrix} \hat{u}(\boldsymbol{\xi}, \boldsymbol{\eta}_{-1})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

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

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

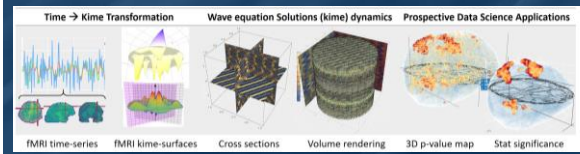
# A *Spacekime* Solution to Wave Equation

- ❑ 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)

# Kime transforms → PDEs → AI

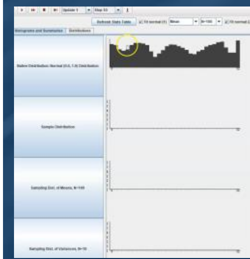


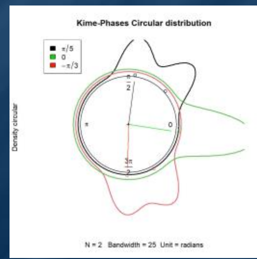

Wang et al., 2022 | Dinov & Velev (2021)

# Hidden Variable Theory & Random Sampling

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

http://wiki.stat.ucla.edu/socr/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)

# *(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 $X$, $f(\boldsymbol{x}, t) \in L^1(X, \mu)$ be an integrable function (e.g., velocity of a particle), and $T: X \to X$ 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, $\bar{f} = E_t(f) = \int_{\mathbb{R}^3} f(\boldsymbol{x}, t) d\mu_x$, will be equal to the average $f$ of just one particle ($\boldsymbol{x}$) over the entire time span,

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

The spatial probability measure is denoted by $\mu_x$ 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**:

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

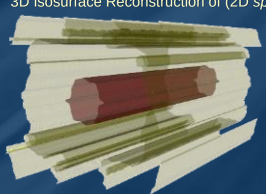
Dinov & Velev (2021)

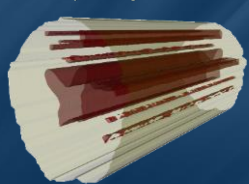
# Mathematical-Physics ⟹ Data Science & AI

| Physics | Data/Neuro 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, $\boldsymbol{O} = \{X, Y\}$ |
| 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 |
| … | … |

# *Spacekime* Analytics: fMRI Example

- 3D Isosurface Reconstruction of (2D *space* × 1D *time*) fMRI signal





**Spacetime**: Reconstruction using trivial phase-angle; kime=time=(magnitude, 0)

**Spacekime**: Reconstruction using correct kime=(magnitude, phase)

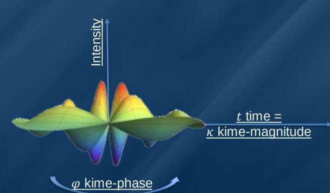
**3D pseudo-spacetime reconstruction:**

$$f = \hat{h}\,(\underbrace{x_1, x_2}_{space}, \underbrace{t}_{time})$$

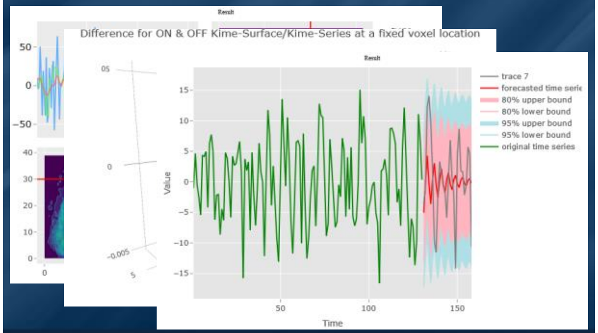
# *Spacekime* Analytics: Kime-series = Surfaces (not curves)

In the 5D spacekime manifold, time-series curves extend to kime-series, i.e., surfaces parameterized by kime-magnitude ($t$) and the kime-phase ($\varphi$)

Kime-phase aggregating operators that can be used to transform standard time-series curves to spacekime kime-surfaces, which can be modeled, interpreted, and predicted using advanced spacekime analytics



## Spacetime Time-series ⇒ Spacekime Kimesurfaces ⇒ TLM

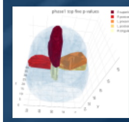



Zhang et al., 2022 | Dinov & Velev (2021)

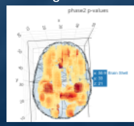
## Tensor-based Linear Modeling of fMRI

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

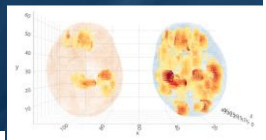
The dimensions of the tensor $Y$ are $160 \times \overbrace{a \times b \times c}^{\text{ROI b-box}}$, where the tensor elements represent the response variable $Y[t, x, y, z]$, i.e., fMRI intensity. For fMRI magnitude (real-valued signal), the design tensor $X$ dimensions are: $\underbrace{160}_{\text{time}} \times \underbrace{4}_{\text{effects}} \times \underbrace{1}_{\mathbb{R}}$.



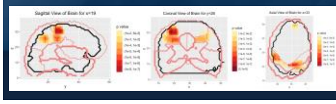
Step 1: ROI analysis



Step 2: Voxel analysis



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



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

## 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 $\gamma$ and the prior distribution, $p(\gamma \mid \varphi)$, of the parameter $\gamma$ before the data $X$ is observed, $\varphi = \text{phase aggregator}$.
- Assume that the hyperparameter (vector) $\varphi$, 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 computed via Laplace transform.
- Let the posterior distribution of the parameter $\gamma$ given the observed data $X = \{x_{i_o}\}$ be $p(\gamma | X, \varphi')$ and the process parameter distribution of the kime-phase hyperparameter vector $\varphi$ be $\gamma \sim p(\gamma \mid \varphi)$.

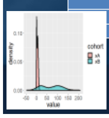
## Bayesian Inference Simulation

- Simulation example using 2 random samples drawn from mixture distributions each of $n_A = n_B = 10K$ observations:
  - $\{X_{A,i}\}_{i=1}^{n_A}$, where $X_{A,i} = 0.3U_i + 0.7V_i$, $U_i \sim N(0,1)$ and $V_i \sim N(5,3)$, and
  - $\{X_{B,i}\}_{i=1}^{n_B}$, 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:
  - Transform all four cohorts into Fourier k-space,
  - Iteratively randomly sample single observations from the (training) cohort $C$,
  - Reconstruct the data into spacetime using a single kime-magnitude value and alternative kime-phase estimates derived from cohorts $B$, $C$, and $D$, and
  - 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.

## 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-phases (in this case, kime-phases derived from cohorts $B$, $C$, and $D$).

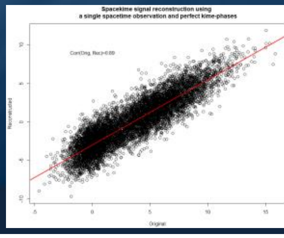
| | **Spacetime** | **Spacekime Reconstructions (single kime-magnitude)** | | |
|---|---|---|---|---|
| **Summaries** | **(A)** **Original** | $(B)$ Phase=Diff. Process | $(C)$ Phase=True | $(D)$ Phase=Independent |
| **Min** | -2.38798 | -3.798440 | -2.98116 | -2.69808 |
| **1st Quartile** | -0.89359 | -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** |




## 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:

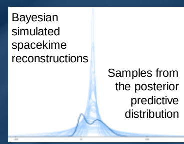
$X_A = 0.3U + 0.7V$, where $U \sim N(0,1)$ and $V \sim N(5,3)$

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

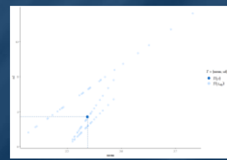
Relations between the empirical data distribution (dark blue) and samples from the posterior predictive distribution, representing Bayesian simulated spacekime reconstructions (**light-blue**). The derived Bayesian estimates do not perfectly match the empirical distribution of the simulated data, yet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

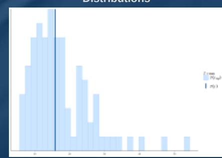
# Bayesian Inference Simulation



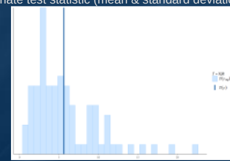

**Distributions**



Bivariate test statistic (mean & standard deviation)



**Test statistic (maximum)**



Test statistic (inter-quartile range, IQR)

Relations between the empirical data distribution (**dark blue**) and samples from the posterior predictive distribution, Bayesian simulated spacekime reconstructions (**light-blue**).

# *Spacekime* Analytics: **Demos**

- Tutorials
  - https://TCIU.predictive.space
  - https://SpaceKime.org
- R Package
  - https://cran.rstudio.com/web/packages/TCIU
- GitHub
  - https://github.com/SOCR/TCIU

# Acknowledgments

Slides Online: "SOCR News"

**Funding**

NIH: UL1TR002240, R01CA233487, R01MH121079, R01MH126137, T32GM141746

NSF: 1916425, 1734853, 1636840, 1416953, 0716055, 1023115

**Collaborators**

- SOCR: **Milen Velev, Yueyang Shen, Daxuan Deng, Zijing Li, Yongkai Qiu, Zhe Yin, Yufei Yang, Yuxin Wang, Rongqian Zhang, Yuyao Liu, Yupeng Zhang, Yunjie Guo, Simeone Marino**
- UMich MIDAS/MCAIM Centers: Lydia Bieri, Kayvan Najarian, Chris Monk, Issam El Naqa, HV Jagadish, Brian Athey

https://SOCR.umich.edu