

**International Conference:  
Differential Equations for Data Science 2023  
(DEDS2023)**

Monday, February 20 — Wednesday, February 22, 2023  
Online (Zoom)

<https://scheme.hn/deds2023/>

Organizers:

Hayato Chiba (Tohoku University, Japan)  
Yoshikazu Giga (The University of Tokyo, Japan)  
Lyudmila Grigoryeva (University of St. Gallen, Switzerland)  
Boumediene Hamzi (California Institute of Technology, US)  
Masato Kimura (Kanazawa University, Japan)  
Hiroshi Kokubu (Kyoto University, Japan)  
Kohei Nakajima (The University of Tokyo, Japan)  
Hirofumi Notsu (Kanazawa University, Japan, Chair)  
Juan-Pablo Ortega (Nanyang Technological University, Singapore)

Registration:

Registration link for Zoom meeting (shortened): <https://tinyurl.com/2hu4xkut>  
After registering, you will receive a confirmation email containing information about joining the meeting. (Once you register from the link above, you are able to join all the sessions.) Note that the Zoom meeting will be available from 30 minutes before each session start time and that the maximum number of registrants is 500.

Supports:

MIRS, Kanazawa University and JST CREST, JPMJCR2014

# Program

\* All lectures are given by invited speakers.

## Monday, February 20

16:55 – 17:00 **Opening**

Session 1 (5 talks) Chair: H. Notsu: JST 17:00 – 20:05

(= UTC 08:00 – 11:05 = CET 09:00 – 12:05 = PST 00:00 – 03:05)

17:00 – 17:45 Houman Owhadi (California Institute of Technology, US)

*On solving/learning nonlinear PDEs with GPs*

17:45 – 17:55 **Break**

17:55 – 18:40 Pantelis R. Vlachas (ETH Zurich/AI2C Technologies, Switzerland)

*Learning and forecasting the effective dynamics of complex systems across scales*

18:40 – 18:50 **Break**

18:50 – 19:15 Yasamin Jalalian (California Institute of Technology, US)

*Forecasting Hamiltonian dynamics with computational graph completion*

19:15 – 19:40 Jonghyeon Lee (California Institute of Technology, US)

*Forecasting dynamical systems from irregularly-sampled data with kernel methods*

19:40 – 20:05 Saad Qadeer (Pacific Northwest National Laboratory, US)

*Machine-learning-based spectral methods for partial differential equations*

## Tuesday, February 21

Session 2 (6 talks) Chair: TBD: JST 17:00 – 20:10

(= UTC 08:00 – 11:10 = CET 09:00 – 12:10 = PST 00:00 – 03:10)

17:00 – 17:45 Boumediene Hamzi (California Institute of Technology, US)

*Kernel Flows and Kernel Mode Decomposition for learning dynamical systems from data*

- 17:45 – 17:55 Break
- 17:55 – 18:20 Jinqiao Duan (Illinois Institute of Technology, US)  
*Nonlocal Kramers–Moyal formulas and data science*
- 18:20 – 18:45 James V. Koch (Pacific Northwest National Laboratory, US)  
*Structural inference of networked dynamical systems*
- 18:45 – 18:55 Break
- 18:55 – 19:20 Jun Okamoto (Kyoto University, Japan)  
*On a singular limit of the Kobayashi–Warren–Carter energy*
- 19:20 – 19:45 Rui Carvalho (Durham University, UK)  
*Automatically identifying dynamical systems from data*
- 19:45 – 20:10 Lingkai Kong (Georgia Institute of Technology, US)  
*Momentum Stiefel optimizer, with applications to suitably-orthogonal attention, and optimal transport*

### Wednesday, February 22

Session 3 (6 talks) Chair: TBD: Session 3: JST 17:00 – 20:10

(= UTC 08:00 – 11:10 = CET 09:00 – 12:10 = PST 00:00 – 03:10)

- 17:00 – 17:45 Lyudmila Grigoryeva (University of St. Gallen, Switzerland, and University of Warwick, UK)  
*Reservoir kernels and Volterra series*
- 17:45 – 17:55 Break
- 17:55 – 18:20 Daiying Yin (Nanyang Technological University, Singapore)  
*Learnability of linear port-Hamiltonian systems*
- 18:20 – 18:45 Giovanni Ballarin (University of Mannheim, Germany)  
*Memory of recurrent networks: Do we compute it right?*
- 18:45 – 18:55 Break

18:55 – 19:20 Satoshi Sunada (Kanazawa University/JST PRESTO, Japan)

*Neural delay differential equations and their physical implementations*

19:20 – 19:45 Andrew Flynn (University College Cork, Ireland)

*From seeing double to modelling seizure dynamics with multifunctional reservoir computers*

19:45 – 20:10 Tina Mai (Duy Tan University, Vietnam, and Texas A&M University, US)

*Prediction of numerical upscaling for Richards equation using deep learning*

20:10 – 20:15 Closing

# Abstract

T. = Title, A. = Abstract.

**1. Houman Owhadi (California Institute of Technology, US)**

T. On solving/learning nonlinear PDEs with GPs

A. We present a simple, rigorous, and unified framework for solving and learning arbitrary nonlinear PDEs with Gaussian Processes (GPs). The proposed approach: (1) provides a natural generalization of collocation kernel methods to nonlinear PDEs and Inverse Problems, (2) has guaranteed convergence for a very general class of PDEs, and (3) comes with a Bayesian interpretation compatible with a UQ pipeline. It inherits (1) the a priori error bounds of kernel interpolation methods and (2) the (near-linear) state-of-the-art computational complexity of linear solvers for dense kernel matrices. Its generalization to high-dimensional and parametric PDEs comes with error bounds exhibiting a tradeoff between dimensionality and regularity (the curse of dimensionality disappears when the problem is sufficiently regular). Its formulation can be interpreted and generalized as an extension of Gaussian Process Regression from the approximation of input/output functions to the completion of arbitrary computational graphs representing dependencies between multiple known and unknown functions and variables.

**2. Pantelis R. Vlachas (ETH Zurich/AI2C Technologies, Switzerland)**

T. Learning and forecasting the effective dynamics of complex systems across scales

A. Predictive simulations of complex systems are essential for applications ranging from weather forecasting to drug design. The veracity of these predictions hinges on their capacity to capture the effective system dynamics. Massively parallel simulations predict the system dynamics by resolving all spatiotemporal scales, often at a cost that prevents experimentation while their findings may not allow for generalisation. On the other hand reduced order models are fast but limited by the frequently adopted linearization of the system dynamics and/or the utilization of heuristic closures. Here we present a novel systematic framework that bridges large scale simulations and reduced order models to Learn the Effective Dynamics (LED) of diverse complex systems. The framework forms algorithmic alloys between non-linear machine learning algorithms and the Equation-Free approach

for modeling complex systems. LED deploys autoencoders to formulate a mapping between fine and coarse-grained representations and evolves the latent space dynamics using recurrent neural networks. The algorithm is validated on benchmark problems and we find that it outperforms state of the art reduced order models in terms of predictability and large scale simulations in terms of cost. LED is applicable to systems ranging from chemistry to fluid mechanics and reduces the computational effort by up to two orders of magnitude while maintaining the prediction accuracy of the full system dynamics. We argue that LED provides a novel potent modality for the accurate prediction of complex systems.

**3. Yasamin Jalalian (California Institute of Technology, US)**

- T. Forecasting Hamiltonian dynamics with computational graph completion
- A. Hamiltonian dynamics describe many different physical systems and have wide range of applications from classical to statistical and quantum mechanics. As such, data-driven simulations of Hamiltonian systems are important tools for solving many scientific and engineering problems and have thus been more widely explored during recent years. In this work, we combine the newly developed framework for computational graph completion (CGC) with numerical techniques for data-adaptive kernel regression to interpolate and forecast Hamiltonian systems in a data-driven way. The CGC framework allows us to characterize the dependencies between the unknowns of the system and approximate them by imposing Gaussian priors and computing MAP estimators given the available data. We demonstrate that our method is both accurate and data-efficient on a variety of physical problems including mass-spring systems, a nonlinear pendulum, and the Hénon–Heiles system.

**4. Jonghyeon Lee (California Institute of Technology, US)**

- T. Forecasting dynamical systems from irregularly-sampled data with kernel methods
- A. A highly efficient way to predict the future of a dynamical system is to interpolate its vector field with a kernel, where the kernel parameters are learned with an algorithm called Kernel Flows (KF), which uses gradient-based optimization to learn a kernel. However, the classical KF algorithm fails if the observed time series is not regularly sampled in time. In our paper, we solve this problem with a generalization of the flow map of the

dynamical system by incorporating time differences between observations in the KF data-adapted kernels; this simple modification leads to a greater forecasting accuracy upon comparison with the original KF algorithm.

**5. Saad Qadeer (Pacific Northwest National Laboratory, US)**

- T. Machine-learning-based spectral methods for partial differential equations
- A. A major obstacle in the deployment of spectral methods is the choice of appropriate bases for trial and test spaces. If chosen suitably, these basis functions lead invariably to well-posed discretized problems and well-conditioned linear systems, while the resulting approximate solutions are provably high-order accurate. However, barring domain decomposition approaches, devising such functions for arbitrary geometries from scratch is a hugely challenging task. Fortunately, recently developed operator learning approaches for approximating solution operators, e.g., DeepONets, Fourier Neural Operators, etc., suggest a highly promising route for generating machine-learned basis functions. In this talk, we propose a Galerkin approach for time-dependent PDEs that is powered by basis functions gleaned from the DeepONet architecture. We shall outline our procedure for obtaining these basis functions and detail their many favourable properties. Next, we shall present the results of numerical tests for various problems, including advection, advection-diffusion, viscous and inviscid Burgers', Korteweg-De Vries, and Kuramoto-Sivashinsky equations. Finally, we will identify potential obstacles in the course of generalization to higher dimensions and suggest possible remedies.

**6. Boumediene Hamzi (California Institute of Technology, US)**

- T. Kernel Flows and Kernel Mode Decomposition for learning dynamical systems from data
- A. Regressing the vector field of a dynamical system from a finite number of observed states is a natural way to learn surrogate models for such systems. We present variants of the method of Kernel Flows as simple approaches for learning the kernel that appear in the emulators we use in our work. First, we will talk about the method of parametric and nonparametric kernel flows for learning chaotic dynamical systems. We'll also talk about learning dynamical systems from irregularly-sampled time series as well as from partial observations. We will also introduce the method of Sparse Kernel Flows and apply it to learn 132 chaotic dynamical systems. Finally,

we extend the method of Kernel Mode Decomposition to design kernels in view of detecting critical transitions in some fast-slow random dynamical systems.

**7. Jinqiao Duan (Illinois Institute of Technology, US)**

T. Nonlocal Kramers–Moyal formulas and data science

A. Dynamical systems in engineering and science are usually under random fluctuations (either Gaussian or non-Gaussian noise). Observational, experimental and simulation data for such systems are noisy and abundant. The governing laws for complex dynamical systems are sometimes not known or not completely known.

This presentation is about extracting stochastic governing laws from noisy data for dynamical systems under non-Gaussian fluctuations, by non-local Kramers–Moyal formulas. I will also compare this approach with the (local) Kramers–Moyal formulas in classical case when noisy fluctuations are Gaussian.

This is a joint work with Yang Li and Yubin Lu.

**8. James V. Koch (Pacific Northwest National Laboratory, US)**

T. Structural inference of networked dynamical systems

A. Data-driven modeling of dynamical systems has experienced a surge in the method development in concert with advances in machine learning and artificial intelligence. In this work, we restrict our scope specifically to the problem of data-driven modeling of networked systems; a challenging problem in which one needs to elicit not only the intrinsic physics of individual nodes of the network, but also what nodes talk to whom and how that communication influences nodal behaviors. The implication of such methodology is far-reaching—one can begin to infer properties of networked systems with respect to network topology and/or external perturbations, answering hypotheticals such as “what would happen if we remove this node?”

**9. Jun Okamoto (Kyoto University, Japan)**

T. On a singular limit of the Kobayashi–Warren–Carter energy

A. We consider the singular limit problem of a single-well Modica–Mortola energy and the Kobayashi–Warren–Carter energy. In this study, we introduce



a finer topology of sliced graph convergence of functions into the function space and derive the singular limit of a single-well Modica–Mortola energy and the Kobayashi–Warren–Carter energy energies in the sense of Gamma-convergence. The energy functional obtained as this singular limit is also shown to have the remarkable property of a minimizing function that is concave concerning the strength of jumps of a function.

**10. Rui Carvalho (Durham University, UK)**

- T. Automatically identifying dynamical systems from data
- A. Discovering governing equations from data provides a clearer understanding of the world around us. Scientists have recently deployed machine learning to develop prediction models representing the expansion of many natural occurrences over time. Here, we present our automatic regression for governing equations (ARGOS) method to extract dynamical systems from noisy data. We expand several linear and nonlinear examples to develop a systematic comparison between the identification performance of ARGOS and the recently proposed SINDy with AIC. Our results show that ARGOS demonstrates a higher identification probability for systems of ordinary differential equations contaminated with state measurement noise.

**11. Ling kai Kong (Georgia Institute of Technology, US)**

- T. Momentum Stiefel optimizer, with applications to suitably-orthogonal attention, and optimal transport
- A. This talk will report a construction of momentum-accelerated gradient descent algorithms on Riemannian manifolds, focusing on a particular case known as Stiefel manifold. The treatment will be based on, firstly, the design of continuous-time optimization dynamics on the manifold, and then a thoughtful time-discretization that preserves all geometric structures. Since Stiefel manifold corresponds to matrices that satisfy orthogonality constraint, two practical applications will also be described: (1) we markedly improved the performance of trained-from-scratch Vision Transformer by appropriately placing orthogonality into its self-attention mechanism, and (2) our optimizer also makes the useful notion of Projection Robust Wasserstein Distance for high-dim. optimal transport even more effective.

**12. Lyudmila Grigoryeva (University of St. Gallen, Switzerland, and University of Warwick, UK)**

T. Reservoir kernels and Volterra series

A. A universal kernel is constructed whose sections approximate any causal and time-invariant filter in the fading memory category with inputs and outputs in a finite-dimensional Euclidean space. This kernel is built using the reservoir functional associated with a state-space representation of the Volterra series expansion available for any analytic fading memory filter. It is hence called the Volterra reservoir kernel. Even though the state-space representation and the corresponding reservoir feature map are defined on an infinite-dimensional tensor algebra space, the kernel map is characterized by explicit recursions that are readily computable for specific data sets when employed in estimation problems using the representer theorem. We showcase the performance of the Volterra reservoir kernel in a popular data science application in relation to bitcoin price prediction.

Paper: <https://arxiv.org/abs/2212.14641>

Joint work with Lukas Gonon (Imperial College London) and Juan-Pablo Ortega (NTU, Singapore)

**13. Daiying Yin (Nanyang Technological University, Singapore)**

T. Learnability of linear port-Hamiltonian systems

A. A well-specified parametrization for single-input/single-output (SISO) linear port-Hamiltonian systems amenable to structure-preserving supervised learning is provided. The construction is based on normal form controllable and observable Hamiltonian representations for those systems, which reveal fundamental relationships between classical notions in control theory and crucial properties in the machine learning context, like structure-preservation and expressive power. More explicitly, it is shown that the equivalence classes of system automorphisms of linear port-Hamiltonian systems can be explicitly identified with a smooth manifold endowed with global Euclidean coordinates, which allows concluding that the parameter complexity necessary for the replication of the dynamics is only  $\mathcal{O}(n)$  and not  $\mathcal{O}(n^2)$ , as suggested by the standard parametrization of these systems. Furthermore, we show that linear port-Hamiltonian systems can be learned while remaining agnostic about the dimension of the underlying data-generating system. Numerical experiments show that this methodology can be used to efficiently estimate linear port-Hamiltonian systems

out of input-output realizations, making the contributions in this paper the first example of a structure-preserving machine learning paradigm for linear port-Hamiltonian systems based on explicit representations of this model category.

**14. Giovanni Ballarin (University of Mannheim, Germany)**

- T. Memory of recurrent networks: Do we compute it right?
- A. Numerical evaluations of memory capacity (MC) of linear recurrent networks reported in the literature often contradict known theoretical bounds. In this paper, we study the case of linear echo state networks, for which total memory capacity is proven to be equal to the range of the Kalman controllability matrix. We shed light on various issues that lead to inaccurate numerical estimations of the memory. We investigate and explain in detail the consequences of neglecting them and we prove that when the Krylov structure of linear MC is ignored, it introduces a “memory gap” between its theoretical and empirical values. As a solution, we develop robust numerical approaches by exploiting a neutrality result of MC with respect to the input mask matrix. Simulations show that memory curves which fully agree with the theory are recovered using the proposed methods.

Joint work with Lyudmila Grigoryeva and Juan-Pablo Ortega.

**15. Satoshi Sunada (Kanazawa University/JST PRESTO, Japan)**

- T. Neural delay differential equations and their physical implementations
- A. Recent work has revealed an interesting connection between deep neural networks and dynamical systems. In this context, the layer-to-layer information propagation in neural nets can be expressed as the time evolution of dynamical systems. The training of deep neural nets has an association with optimal control of dynamical systems. Here, based on our previous work [1], we introduce a new class of neural ordinary differential equations with time delay and its training scheme based on an optimal control theory. We show that an optimally-controlled delay system can perform pattern recognition only with a few control signals and a single node [1], in contrast to standard deep neural nets with a huge number of weight parameters and neurons. The feature of controlled delay equations practically allows for its simple physical implementations. In this talk, we will introduce an optoelectronic neural delay system [2] and a new class of training strategies

without back-propagation, which is based on direct feedback alignment (DFA) with correlated randomness [3].

- [1] G. Furuhata, T. Niiyama, and S. Sunada, “Physical deep learning based on optimally controlled dynamical systems,” *Phys. Rev. Applied* 15, 034092 (2021).
- [2] R. Nogami, K. Kanno, S. Sunada, and A. Uchida, “Experimental demonstration of physical deep learning based on optimal control using optoelectronic delay system,” *Proc of NOLTA 2022*, B3L-C-02 (2022).
- [3] In preparation.

**16. Andrew Flynn (University College Cork, Ireland)**

- T. From seeing double to modelling seizure dynamics with multifunctional reservoir computers
- A. In the pursuit of developing artificially intelligent systems there is much to be gained from dually integrating further physiological features of biological neural networks and knowledge of dynamical systems into machine learning environments. In this talk such a two-armed approach is employed in order to translate ‘multifunctionality’ from biological to artificial neural networks via the reservoir computing machine learning paradigm. Multifunctionality describes the ability of a single neural network to perform a multitude of mutually exclusive tasks by exploiting a form of multistability. The dynamics of how a reservoir computer achieves multifunctionality when tasked with solving the ‘seeing double’ problem are presented. These results help to identify many new application areas for reservoir computers which are also explored in this talk including, data-driven modelling of multistability, generating chaotic itinerancy for memory recall, and reconstructing dynamical transitions present in the epileptic brain.

**17. Tina Mai (Duy Tan University, Vietnam, and Texas A&M University, US)**

- T. Prediction of numerical upscaling for Richards equation using deep learning
- A. In [Sergei Stepanov, Denis Spiridonov, and Tina Mai. Prediction of numerical homogenization using deep learning for the Richards equation. *Journal of Computational and Applied Mathematics*, 424:114980, 2023. <https://doi.org/10.1016/j.cam.2022.114980>], for an unsaturated flow in the form of nonlinear Richards equation over heterogeneous media, we build

a new coarse-scale approximation scheme based on numerical homogenization. Using deep neural networks (DNNs), this strategy provides frequent and rapid estimates of macroscopic parameters. To be more precise, during training a neural network, we employ a training set of random permeability realizations and correspondingly computed macroscopic targets (effective permeability tensor, homogenized stiffness matrix, and right-hand side vector). Our proposed deep learning approach, which constructs nonlinear maps between such permeability fields and macroscopic properties, is novel in that it treats the nonlinearity of Richards equation in the predicted coarse-scale homogenized stiffness matrix. Numerous numerical experiments on problems involving two-dimensional models show how well this method predicts the macroscopic features and hence solutions.