“Big data analytical methods for complex systems”

Wrocław, October 6-7, 2022

ABSTRACTS

CASUS CENTER FOR ADVANCED SYSTEMS UNDERSTANDING
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
Uniwersytet Wrocławski

From biology and models of carcinogenesis to mathematical theorems

Grzegorz Karch[*]

Mathematical Institute, University of Wroclaw

I shall describe my adventures, as a pure mathematician, in the world of theoretical biology and theoretical medicine which are developed in Heidelberg University: in Institute of Applied Mathematics and in BIOQUANT Center. Together with

- Prof. Dr. Anna Marciniak-Czochra, Heidelberg University, Germany
- Prof. Dr. Kanako Suzuki, Ibaraki University, Japan
- Dr. Szymon Cygan, University of Wroclaw, Poland

we studied certain models of early carcinogenesis, that is, certain reaction-diffusion systems where ordinary differential equations are coupled with one reaction-diffusion equation. In our mathematical works, we proved that such systems may have regular (i.e. sufficiently smooth) stationary solutions, however, all of them are unstable. Then, we showed that solutions for such *reaction-diffusion-ODE* systems behave in a singular way for large values of time which means that they may be infinite in finite time or they may converge towards discontinuous steady states.

Some our publications are listed below.

[1] Cygan, Szymon; Marciniak-Czochra, Anna; Karch, Grzegorz; Suzuki, Kanako; Instability of all regular stationary solutions to reaction-diffusion-ODE systems. J. Differential Equations 337 (2022), 460–482.

[2] Cygan, Szymon; Marciniak-Czochra, Anna; Karch, Grzegorz; Suzuki, Kanako; Stable discontinuous stationary solutions to reaction-diffusion-ODE systems. (2022), 1-30. arXiv:2111.01214 [math.AP]

[3] Cygan, Szymon; Marciniak-Czochra, Anna; Karch, Grzegorz; Discontinuous stationary solutions to certain reaction-diffusion systems. Partial Differ. Equ. Appl. 3 (2022), no. 4, Paper No. 49, 15 pp.

[4] Marciniak-Czochra, Anna; Karch, Grzegorz; Suzuki, Kanako; Instability of Turing patterns in reaction-diffusion-ODE systems. J. Math. Biol. 74 (2017), no. 3, 583–618.

[*] Email: Grzegorz.Karch@uwr.edu.pl

Statistical Learning at UWr

Malgorzata Bogdan

Institute of Mathematics, University of Wroclaw, Wroclaw, Poland

In this talk we will give a short overview of the Statistical Learning research performed at the University of Wroclaw. The main emphasis will be devoted to the research in Large Data Analysis, where the main goal is the development of efficient computer algorithms with mathematical guarantees concerning their statistical properties. We will present the directions of the theoretical and implementation work and point at some applications in medicine, genetics, finance, and astronomy. This research is performed in a broad collaboration with top statisticians and experts in specific fields from Europe and USA.

Network Optimization

Marcin Bieńkowski[1]

[1] *Institute of Computer Science, University of Wroclaw, Wroclaw, Poland*

In the last 20 years, I have been doing theoretical research in combinatorial optimization, solving problems motivated by applications from networking and logistics. In this talk, I highlight some attempts to make this research applicable in real-life scenarios.

[1] Email: marcin.bienkowski@cs.uni.wroc.pl

Discovering the structure: unsupervised methods in speech recognition

Paweł Rychlikowski[*1]

[1] *Institute of Computer Science, University of Wroclaw, Wroclaw, Poland*

In this talk we consider questions related to unsupervised or semi-supervised methods in natural language processing, with the main focus on speech recognition. We briefly introduce such (quite) modern concepts like Word2Vec, BERT, Contrastive Predictive Coding, Wave2vec-U, and our ongoing research in this area. We also discuss the fundamental question: how much of the language can be reconstructed when we have only recorded voice, without any connections to a text? Can we say what is a word, and what is not? Can we discover the meaning of spoken words? Can we say that something is a grammatically correct sentence? These questions are related to the idea of ZeroSpeech Challenge in which we have to imitate the 'infant strategy' of language acquisition rather than following standard machine learning approach with thousands of labeled recordings.

[*] Email: prych@cs.uni.wroc.pl

Animal movement research as a cross-cutting theme at CASUS

Justin M. Calabrese[*1, 2, 3] , **Inês Silva**[1], **Jesse M. Alston**[1,4], **Chris H. Fleming**[2]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*
[2] *Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany*
[3] *Department of Biology, University of Maryland, College Park, USA*
[4] *School of Natural Resources and the Environment, University of Arizona, Tucson, USA*

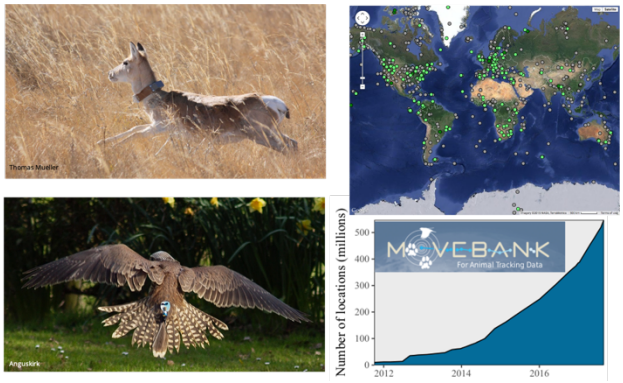


Figure 1: Technological advances have fueled an explosion of high-resolution animal tracking data

The movement of animals through landscapes worldwide shapes biodiversity patterns, influences disease transmission, and governs how humans and wildlife interact. High resolution animal tracking data have transformed our ability to understand when, where, how, and why animals move. However, these data come with formidable statistical challenges including strong autocorrelation and context-dependent location errors and fix rates. Overcoming these hurdles requires an interdisciplinary effort that combines ecology, physics, geostatistics, signal processing, and computer science.

In this talk, I detail ongoing work at CASUS in animal movement research, covering statistical methods and software development as well as applications in ecology, wildlife management, and autonomous vehicles research. I also highlight the role that aggregated, multispecies tracking datasets play in understanding animal movement and its consequences at the global scale. Finally, I discuss future directions for this research program, outlining potential points of collaboration with researchers coming from different disciplines.

* Email: j.calabrese@hzdr.de

Machine Learning in Biomedical Images to Study Infection and Disease

Artur Yakimovich[*1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*

Recent advances in Machine Learning (ML) and Deep Learning (DL) are revolutionizing our abilities to analyze biomedical images and deepen our understanding of infection and disease. Among other host-pathogen interactions may be readily deciphered from microscopy data using convolutional neural networks (CNN). ML/DL algorithms may allow unambiguous scoring of virus-infected and uninfected cells in absence of specific labeling. Furthermore, accompanied by interpretability approaches, the ability of CNNs to learn representations, without explicit feature engineering, may allow uncovering yet unknown phenotypes in microscopy. One such example is our recent tandem segmentation-classification algorithm aimed to uncover morphological markers of Caenorhabditis elegans lifespan and motility. Taken together these novel approaches may facilitate novel discoveries in Infection and Disease Biology.

[*] Email: a.yakimovich@hzdr.de

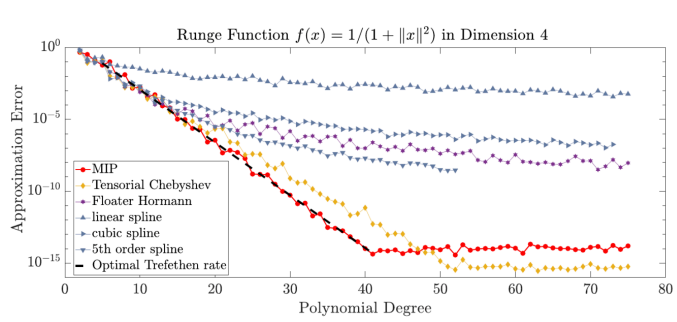
minterpy - Multivariate Interpolation in Python

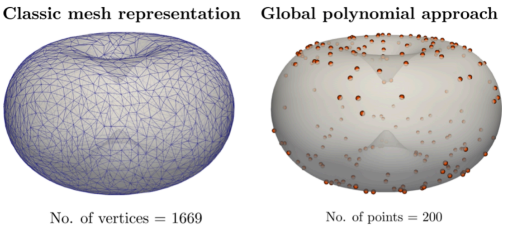
Michael Hecht, Janina Schreiber, Damar Wicaksono

CASUS - Center for Advanced Systems Understanding,
Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany

Many challenges arising in the fields of computational science and engineering rely on solving interpolation tasks of highly-varying sparse and scattered data. The tasks include surrogate modeling, sparse data regression, global black box optimization, model inference, as well as solutions for partial differential equations (PDE) on complex geometries.

Interpolation tasks in multi-dimensional space typically suffer from the *curse of dimensionality* in which the computational cost of interpolation scales exponentially with the number of dimensions. The open-source Python package *minterpy,* developed and maintained by the Hecht-Lab, CASUS, aims to lift the curse of dimensionality from a vast field of interpolation tasks arising across scientific disciplines.





Order of approximation errors

	Classic mesh	Global polynomial
Surface normal	$\sim 10^{-4}$	$\sim 10^{-12}$
Mean curvature	$\sim 10^{-2}$	$\sim 10^{-14}$
Gauss curvature	$\sim 10^{-2}$	$\sim 10^{-10}$

Efficient machine-precision-approximation of strongly varying functions

Global Polynomial implicit parametrisation of 2D surfaces

MINTERPY [1] minterpy Multivariate interpolation, Python (2021) https://github.com/casus/minterpy/releases

The research scope includes (inverse) PDE problems (e.g., for bio-physical cell deformations), interpolation and regression techniques for strongly varying functions (e.g., electron densities of hot matter), model inference and black-box optimization (e.g., for tumor response models), and mathematics in terms of algebraic topology for high dimensional data and general approximation theory.

[1] Hecht, M., and Sbalzarini, I.F. Fast interpolation and Fourier transform in high-dimensional spaces. In Intelligent Computing. Proc. 2018 IEEE Computing Conf., Vol.2, (London, UK, 2018), K. Arai, S. Kapoor, and R. Bhatia, Eds., vol. 857 of Advances in Intelligent Systems and Comput- ing, Springer Nature,pp. 53–75.

[2] M. Hecht et. al, Multivariate Interpolation on Unisolvent Nodes -- Lifting the Curse of Dimensionality' Arxiv, 2020, https://doi.org/10.48550/arXiv.2010.10824

[3] Jannik Michelfeit, 'Interpolation of Hyperparameters for Deep Neural Networks', Diplomarbeit, TU Dresden, 02.11.2020

Physics-informed and data-driven modeling of matter under extreme conditions

Lenz Fiedler[1], **Karan Shah**[1], **Tim Callow**[1], **Kushal Ramakrishna**[1], **Daniel Kotik**[1], **Steve Schmerler**[2], **Attila Cangi**[*1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany*
[2] *Information Services and Computing, Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany*

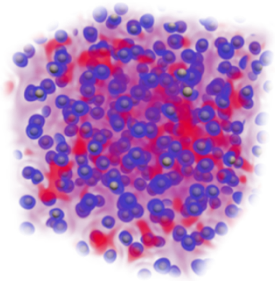


Figure 1: The electronic structure of Aluminum at its melting point.

Understanding the properties of matter under extreme conditions is essential for advancing our fundamental understanding of astrophysical objects and guides the search for exoplanets, it propels the discovery of materials exhibiting novel properties that emerge under high temperatures and pressure, it enables novel technologies such as nuclear fusion, and supports diagnostics of experiments at large-scale brilliant photon sources. While modeling in this challenging research domain has so far relied on first-principles methods [1,2], these turn out to be computationally too expensive for simulations at the required time and length scales. Reduced models, such as average-atom models [3], come at a reduced computational and are useful by connecting atomistic details with hydrodynamics simulations, but they provide less accuracy. Artificial intelligence (AI) has great potential for accelerating electronic structure calculations to hitherto unattainable scales [4]. I will present our recent efforts on accomplishing speeding up Kohn-Sham density functional theory calculations with deep neural networks in terms of our Materials Learning Algorithms framework [5,6] by illustrating results for metals across their melting point. Furthermore, our results towards automated machine-learning save orders of magnitude in computational efforts for finding suitable neural networks and set the stage for large-scale AI-driven investigations [7].

[1] T. Dornheim, A. Cangi, K. Ramakrishna, M. Böhme, S. Tanaka, J. Vorberger, Phys. Rev. Lett. 125, 235001 (2020).
[2] K. Ramakrishna, A. Cangi, T. Dornheim, J. Vorberger, Phys. Rev. B 103, 125118 (2021).
[3] T. J. Callow, E. Kraisler, S. B. Hansen, A. Cangi, Phys. Rev. Research 4, 023055 (2022).
[4] L. Fiedler, K. Shah, M. Bussmann, A. Cangi, Phys. Rev. Materials 6, 040301 (2022).
[5] A. Cangi et al., MALA, https://doi.org/10.5281/zenodo.5557254 (2021).
[6] J. A. Ellis, L. Fiedler, G. A. Popoola, N. A. Modine, J. A. Stephens, A. P. Thompson, A. Cangi, S. Rajamanickam, Phys. Rev. B 104, 035120 (2021).
[7] L. Fiedler, N. Hoffmann, P. Mohammed, G. A. Popoola, T. Yovell, V. Oles, J. A. Ellis, S. Rajamanickam, A. Cangi, arXiv:2202.09186 (2022).

[*] Email: a.cangi@hzdr.de

Frontiers of computational quantum many-body theory

Tobias Dornheim [*1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany*

The fundamental laws necessary for the mathematical treatment of a large part of physics and the whole of chemistry are thus completely known, and the difficulty lies only in the fact that application of these laws leads to equations that are too complex to be solved." Nearly a century has passed, yet the famous quote by Paul Dirac still gets to the heart of many research fields within theoretical physics, quantum chemistry, material science, etc. In this talk, I will show how we can use cutting-edge numerical methods on modern high-performance computing systems to effectively overcome these limitations in many cases. In this way, we get unprecedented insights into quantum many-body systems on the nanoscale going all the way from ultracold atoms like superfluid helium to warm dense matter that occurs within planetary interiors and thermonuclear fusion applications.

[*] Email: t.dornheim@hzdr.de

Future GNSS troposphere remote sensing, next Big Data for weather forecasting?

Witold Rohm[11], **Adam Cegła**[1], **Estera Trzcina**[1], **Natalia Hanna**[2], **Gregor Moeller**[3], **Paweł Hordyniec**[1]

[1] *Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland*
[2] *Department of Geodesy and Geoinformation, Technical University of Vienna, Vienna, Poland*
[3] *Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Zurich, Switzerland*

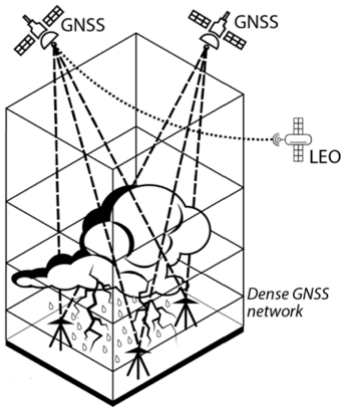


Fig. 1 Concept of ground- and space-based observation GNSS network

Currently, the Global Navigation Satellite Systems observations (ground-based and space-based) are used mainly for positioning, precise orbit determination, time transfer, monitoring of global geodynamic processes, establishing and conserving global and regional reference frames. However, GNSS is also very well suited for improving weather forecasts, which is especially important for short term severe weather predictions (tornadoes, flash floods, tropical cyclones, mesoscale convective systems). The observations from sparsely distributed ground GNSS networks (to provide access to reference frames, support construction, mapping and cadastral applications), space-based low Earth orbiting satellites are being used either as an integrated observation of water vapour or as a profile of bending angle, refractivity, temperature, pressure and humidity. In addition to the high-end receivers used in GNSS networks, low-cost GNSS chips are more and more ubiquitous to hand-held devices, wearables, public and private transport vehicles, including autonomous transport, and drones.

The number of devices that are continuously using GNSS signals for positioning is increasing exponentially, amid autonomous vehicles and drones surge. All these signals provide observations about the state of the troposphere. However, before these observations can be used in weather forecasting, the following challenges have to be solved: 1) the estimation of troposphere is directly correlated with height estimation, therefore stable coordinates are required, 2) collaborative positioning / troposphere estimation might be required, 3) turning the obtained troposphere parameters (delays, refractivity profiles, bending angles) into actionable parameters is not trivial as Numerical Weather Models (NWM) requires forward and adjoint operators to assimilate (correct) model state.

Therefore, within this project we will simulate the observation environment, in terms of buildings, transport modes, satellite signal availability, atmospheric state and based on this information test the input for future numerical weather models.

[1] Email: witold.rohm@upwr.edu.pl

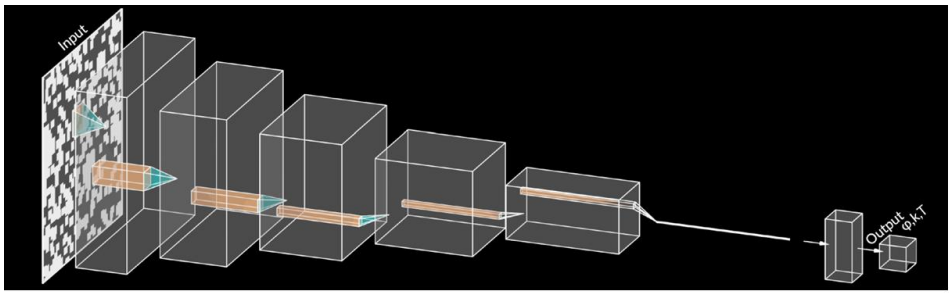
Predicting porous medium properties by deep neural network

Krzysztof M. Graczyk[*1], **Maciej Matyka**[1]

[1] *Institute of Theoretical Physics, University of Wrocław, Wrocław, Poland*

Deep neural networks (DNN) have been utilized in various branches of experimental and theoretical physics. The analyses of the experimental data and optimization of computational systems are the most natural applications. DNNs are also employed in studies of the properties of classical and quantum systems.

In this talk, we shall consider the problem of fluid flow in porous medium. We focus on the predictions of two macroscopic quantities that characterize fluid flow in the porous medium: permeability and tortuosity. For some classes of systems, the permeability and the tortuosity are determined by the system geometry (configuration of obstacles) only.



We will show that the computational system, given by Convolutional Neural Network (CNN), can predict permeability and tortuosity with reasonable accuracy [1]. In our numerical experiments, we considered the 2-dimensional porous medium system. The CNN takes initial configurations of obstacles as input, represented by a picture. As an output, the network gives porosity, permeability, and tortuosity.

The network was trained on the data obtained in the numerical simulations. The fluid flow through a porous medium was simulated with the lattice Boltzmann method. During the talk, we will shortly present the general idea of our approach and discuss the main difficulties and plans for the future.

[1] Krzysztof M. Graczyk and Maciej Matyka, Sci Rep 10, 21488 (2020).

[*] Email: krzysztof.graczyk@uwr.edu.pl

Computational Intelligence in Mining Temporal Data with Hidden Structure Discovering

Piotr Lipiński[1]*,

* *Institute of Computer Science, University of Wrocław, Wrocław, Poland*

Our research concerns data mining from temporal data with hidden data structure discovering. It includes computational intelligence approaches to a few particular applications related to:
- reducing the search space in evolutionary algorithms for high-dimensional optimization problems by discovering a hidden structure of successive populations in the search space and transforming the original search space into a smaller-dimensional one,
- discovering hidden data structure and using it in feature engineering for mining frequent patterns from multidimensional financial ultra-high frequency time series on real-world data from the London Stock Exchange Rebuilt Order Book,
- detecting seasonal patterns in water demand forecasting on real-world data from the water pipeline system of Wroclaw,
- discovering characteristic patterns in saturation curves for selected groups of products for active users in recommender systems.

This work is partially related to a national research project on discovering hidden structure in large datasets (supported by the Polish National Science Centre (NCN) under grant OPUS-18 no. 2019/35/B/ST6/04379).

[1] Brabazon, A., Lipinski, P., Hamill, P., Characterising Order Book Evolution Using Self-Organising Maps, [in] Evolutionary Intelligence, vol. 9, no. 4, Springer, 2016, pp. 167-179.
[2] Michalak, K., Lancucki, A., Lipinski, P., Multiobjective Optimization of Frequent Pattern Models in Ultra-High Frequency Time Series: Stability versus Universality, [in] IEEE CEC, 2016, pp. 3491-3498.
[3] Lipinski, P., Filipiak, P., Rychlikowski, P., Stanczyk, J. Kajewska-Szkudlarek, J., Lomotowski, J., Konieczny, T., Discovering weekly seasonality for water demand prediction using evolutionary algorithms, [in] ACM GECCO, 2017, pp. 33-34.
[4] J. Stanczyk, J. Kajewska-Szkudlarek, P. Lipinski, P. Rychlikowski, Improving short-term water demand forecasting using evolutionary algorithms, Scientific Reports, 2022.

[1] Email: piotr.lipinski@cs.uni.wroc.pl

Nonparametric tests for selected testing problems with applications

Grzegorz Wyłupek

Institute of Mathematics, University of Wrocław,
pl. Grunwaldzki 2, 50-384 Wrocław, Poland

In the talk, we consider several general nonparametric testing problems and briefly discuss their possible solutions. Each problem is motivated by the applications. Therefore, each part of the talk has the same stucture, i.e., motivation, provided by a real data example, the strict mathematical formulation, expressed in terms of the hypotheses testing, and, finally, the solution, being either a kind of a data-driven test or a certain functional of the related empirical process. When possible, we also yield some links to the classical tests. We are going to discuss: the anova [1], one-sided two-sample problem [2], validation of positive quadrant dependence [3], testing poissonity [4], two-sample right-censored model [5], and two-sample paired-data problem [6]. All cases are treated as nonparametrics, which means, that we do not impose any distributional assumptions on the underlying data.

[1] G. Wyłupek, Data-driven *k*-sample tests, *Technometrics*, 52, 107-123, (2010).
[2] T. Ledwina, G. Wyłupek, Two-sample test against one-sided alternatives, *Scandinavian Journal of Statistics*, 39, 358-381, (2012).
[3] T. Ledwina, G. Wyłupek, Validation of positive quadrant dependence, *Insurance: Mathematics and Economics*, 56, 38-47, (2014).
[4] T. Ledwina, G. Wyłupek, On Charlier polynomials in testing Poissonity, *Communications in Statistics – Simulation and Computation*, 46, 1918-1932, (2017).
[5] G. Wyłupek, A permutation test for the two-sample right-censored model, *Annals of the Institute of Statistical Mathematics*, 73, 1037-1061, (2021).
[6] G. Wyłupek, A nonaparametric test for paired data, *under review*, (2022).

A novel semiparametric model for hydrogen deuterium exchange monitored by mass spectrometry data.

Krystyna Grzesiak[1], **Weronika Puchała**[2], **Michał Dadlez**[2], **Małgorzata Bogdan**[1], **Michał Burdukiewicz**[3]

[1] *Faculty of Mathematics and Computer Science, University of Wrocław, Wrocław, Poland*
[2] *Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warszawa, Poland*
[3] *Medical University of Białystok, Białystok, Poland*

The hydrogen-deuterium exchange monitored by mass spectrometry (HDX-MS) is one of the methods for studying the structure of proteins. In principle, the proteins are incubated in heavy water. During the incubation more exposed residues undergo the exchange of amide hydrogens to deuters much faster than others [1]. HDX-MS associates the speed of hydrogen-deuterium exchange with the regional rigidity of a protein. Such stability may be affected by the biological state [2] (e.g., presence of a ligand or lack thereof). Therefore, the changes in the protein's molecular structure caused by the biological state are inferred from the differences in the speed of hydrogen-deuterium exchange.
We propose a novel test based on a mixed semiparametric model and ridge regression that allows for the accurate identification of regions with significantly different exchange speeds at two biological states. To assess its performance, we have compared it with existing HDX-MS data analysis methods. The spline models are not as exposed to the impact of outliers as other model-based tests because of their local characteristics. As distinct from all existing methods they are capable of modeling the deuterium uptake well along the time taking into account its variability what makes them more sensitive in discriminating between different biological states.
We simulated data of 72 peptides at different biological states. The analysis of rejection rate in the pairwise testing procedure at the significance level 0.05 showed that the type I error (false positives) of semiparametric test does not surpass 0.06 while its power (true positives) is greater, when compared to other methods, even about 0.6 in the tough cases when the actual differences are small.

[1] A. Berger and K. Linderstrøm-Lang. Arch. Biochem. Biophys. 69,106–118 (1957).

[2] G. R. Masson, J. E. Burke, N. G. Ahn, G. S. Anand, C. Borchers, S. Brier, G. M. Bou-Assaf, J. R. Engen, S. W. Englander, J. Faber, et al. Nat. Methods 16(7),595–602 (2019).

[1] Email: krygrz11@gmail.com

Regularization methods for gene identification

Rodrigo Cofre Ramirez[1]

[1] *Institute of Mathematics, University of Wrocław, pl. Grunwaldzki 2, 50-384 Wrocław, Poland*

The task of localizing quantitative trait loci (QTL) is usually performed by fitting the linear regression model, with explanatory variables specified by the genotypes at specific marker loci, and then determining which of these loci are associated with the trait values. This approach assumes that the number of QTL is relatively small and may lead to many false discoverie (so called ghost QTL) in the situation when the trait is additionally influenced by a large number of very weak genes (polygenic component). It has been previously shown that this problem can be resolved by supplementing the linear regression model with the random polygenic effects. The benchmark method proposed in the literature was calibrated so as to control the probability of identifying at least one ghost QTL. In this talk we will present a novel approach aimed at controlling the false discovery rate (FDR). Our method relies on the novel extension of the Adaptive Bayesian SLOPE (ABSLOPE, Jiang et al. in 2020), which originally was designed for the classical fixed effects model. The performance of this novel implementation is compared to a benchmark procedure. Our simulations show that our method keeps FDR very close to the nominal levels and can identify more QTL than the benchmark at the cost of a small number of false discoveries..

[1] Email: 327160@uwr.edu.pl

Classification with Imbalanced Data: Predicting Instances of Severe Knee and Back Pain

Maciej Grabias[11]

[1]*Institute of Mathematics, University of Wrocław*
pl. Grunwaldzki 2, 50-384 Wrocław, Poland

Machine learning techniques have found numerous applications in various fields of medicine. However, data for classification tasks coming from medical domains often suffers from the problem of class imbalance, as the prevalence of examined conditions is usually low. In this work, machine learning algorithms involving logistic regression and tree-based ensembles have been applied to a medical surveillance dataset of moderate class imbalance with the goal of identifying minority instances of severe pain and relevant explanatory variables associated with this outcome. Extensive experiments show that all considered models yield similarly poor performance achieving very low sensitivities. As an approach for mitigating the class imbalance issue, cost-sensitive learning strategies have been adopted which brought identical effects in all investigated models and allowed for sensitivity and specificity values to reach comparable levels. Important predictors determined through different feature selection methods are in agreement with the actual contributing factors reported by medical practitioners. Parallels between cost-sensitive strategies and classification threshold manipulation are also discussed.

[1] Email: 325922@uwr.edu.pl

Explaining predictability of human movement trajectories through the sequence matching algorithms

Kamil Smolak[1], **Witold Rohm**[1], **Katarzyna Sila-Nowicka** [1,2,3]

[1] *Institute of Geodesy and Geoinformatics, Wrocław University of Environmental and Life Sciences, Wrocław, Poland*
[2] *School of Environment, The University of Auckland, Auckland, New Zealand*
[3] *Urban Big Data Centre, University of Glasgow, Glasgow, Scotland, UK*

Understanding everyday human movement is a crucial element of sustainable urban systems. Although human mobility seems to be random and spontaneous, multiple studies indicate its high regularity and predictability. Up to this day, many sophisticated machine learning algorithms were successfully employed to predict the movement of individuals. These predictions help to efficiently manage transportation networks and plan cities' development.

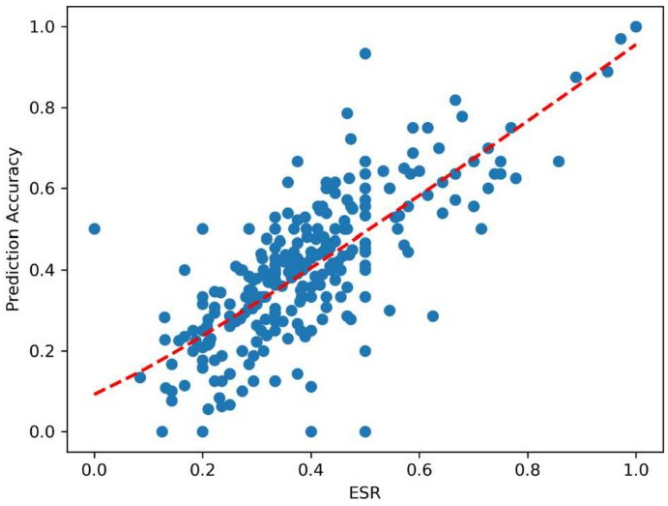


Fig. 1 Comparison of the potential prediction accuracy estimated by the proposed pattern matching-based metric (ESR) and actual accuracy of predictions made by an advanced prediction algorithm.

Human predictability limit is a theory developed to serve as a benchmark for mobility prediction algorithms. It computes an upper bound of prediction accuracy which can be reached by the potentially perfect prediction method. Studies showed that this predictability limit varies significantly when applied to different mobility datasets representing human mobility in various areas. However, the low interpretability of this method impedes the identification of factors determining the predictability of human movement.

This project develops novel pattern matching based-measures which allow for quick estimation of the potential predictability of human movement using an easy-to-interpret algorithm. This method serves as an alternative approach to the complex predictability limit estimation theory. Its output can be analysed to extract highly predictable movement regularities and identify rare travels which are difficult to predict. This, in turn, helps in further development of human movement prediction algorithms. Furthermore, this project demonstrates that predictability limit theory should not be used as a benchmark for mobility prediction algorithms due to multiple discrepancies in the fundamental assumptions of the prediction task and predictability estimation approach.

[1] Email: kamil.smolak@upwr.edu.pl

Locally-informed proposals in Metropolis-Hastings algorithm with applications

Bartosz Chmiela[1]

[1] *Mathematical Institute, University of Wrocław, pl. Grunwaldzki 2/4, 50-384 Wrocław, Poland*

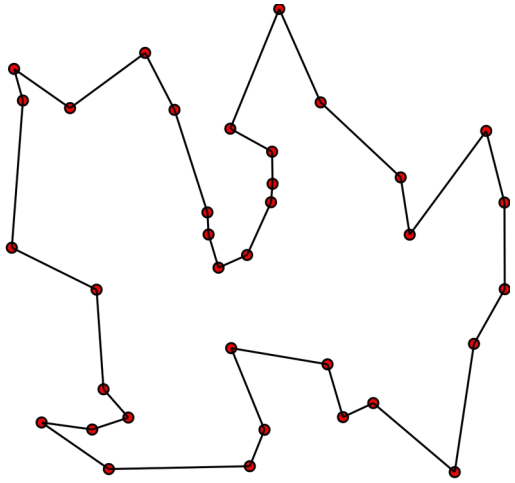


Figure 1: Traveling salesman problem, source: wiki.

The Markov chain Monte Carlo methods (abbrv. MCMC) are a family of algorithms used for approximate sampling from a given probability distribution. They prove very efective when the state space is large. This fact can be used to solve many hard deterministic problems - one of them being traveling salesmen problem, which asks for the shortest route that visits all of the cities exactly once. We will present an application of a relatively new modification of a well known Metropolis-Hasting algorithm (called locally informed proposals) to the aforementioned traveling salesman problem. This approach uses a locally computed distribution, that changes depending on a candidate, at each step of the Metropolis-Hastings algorithm. We will present the implementation of the modified algorithm, experiments based on it, results and a comparison with previous MCMC methods.

Talk based on a master's thesis written under supervision of Paweł Lorek

[1] Email: chmiela.bartosz@gmail.com

Advancing animal movement research

Inês Silva[*1], **Justin M. Calabrese**[1,2,3]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*
[2] *Department of Biology, University of Maryland, College Park, USA*
[3] *Department of Ecological Modelling, Helmholtz Centre for Environmental Research-UFZ, 04318 Leipzig, Germany*

Numerous research questions in ecology relate to how animals interact, live and thrive in a changing world, linking internal state (why move?), navigation capacity (where to move?), and external factors (what environmental factors may affect movement?). Field ecologists collect movement data using animal-borne devices, such as GPS loggers, effectively assessing the animal's behavior, interactions, and the environment in which it lives. The study of animal movement ecology is rapidly shifting as we open new avenues of research, with increased access to modern tracking technologies that enable collection of high-volume, high-resolution movement data for a growing number of animal species worldwide [1].

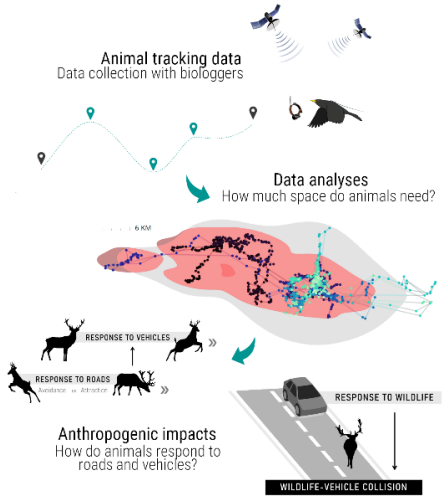


Figure 1. Workflow for animal movement data collection, analyses, and research questions.

At CASUS, we are at the forefront of harnessing the big data revolution to improve ecologists' understanding of animal movement by intersecting mathematics, computer science, and ecology [2]. Work from CASUS scientists has resulted in tools to quantify encounter rates and estimate area requirements [3]. The former has implications for disease transmission, while the latter is required for conservation and wildlife management efforts. Additionally, we are developing a new class of models to study habitat selection by animals that account for biases due to autocorrelation and location-fix failure in animal tracking data [4]. We are also interested in how study design choices in movement ecology influence subsequent analyses, and have created a web-based application to improve experimental study design. In addition, we are exploring how transportation has a rapidly growing impact on wildlife due to vehicle-wildlife collisions, and how the introduction of autonomous vehicles may provide a unique opportunity for implementing mitigation measures [5].

Big movement data will ultimately increase our understanding of threatened species and ecosystems, revealing novel information that will inform conservation and management decisions at the regional and global levels.

[1] Nathan, R., *et al.* Science, 375(6582), eabg1780 (2022).
[2] Calabrese, J. M., *et al.* Methods in Ecology and Evolution, 7(9), 1124-1132 (2016).
[3] Silva, I., *et al.* Methods in Ecology and Evolution, 13(3), 534-544 (2022).
[4] Alston, J. M., *et al.* bioRxiv (2022).
[5] Silva, I., & Calabrese, J. M. ecoevoRxiv (2022).

[*] Email: i.simoes-silva@hzdr.de

Transferability of DFT surrogate models: Temperature and system size

Lenz Fiedler[1,2], **Attila Cangi**[1,2]

[1] *Center for Advanced Systems Understanding (CASUS), D-02826 Görlitz, Germany*
[2] *Helmholtz-Zentrum Dresden-Rossendorf, D-01328 Dresden, Germany*

While Density Functional Theory (DFT) is the most common tool for the investigation of materials under extreme conditions, its scaling behavior with respect to both system size and temperature makes large scale simulations challenging. Yet, progress in this regard would enable accurate modeling of planetary interiors or radiation damage in fusion reactor walls.
One possible route to alleviate these scaling problems is through the use of surrogate models, i.e., machine-learning models. These are trained on DFT data and are able to reproduce DFT observables at comparable accuracy, but negligible computational cost.
In order to actually be useful for such investigations, existing models need to be able to work across length scales and be transferable within desired temperature ranges. Here we show how models based on local mappings of electronic structure information [1], implemented in the Materials Learning Algorithms (MALA) package [2] can be trained on small number of atoms and select temperatures, yet perform accurately when used to make predictions for extended systems within a range of temperatures.

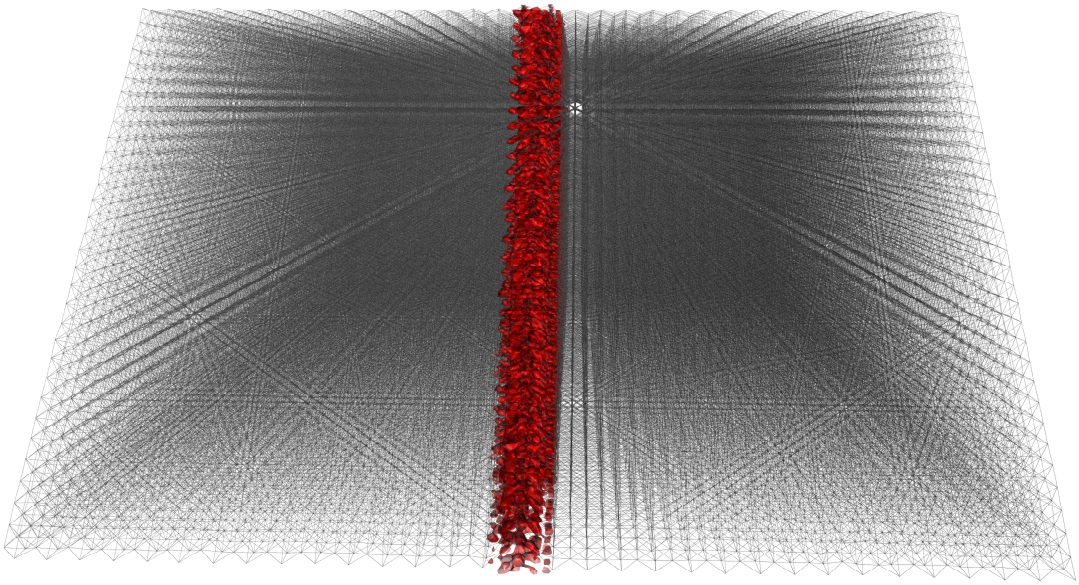


Fig. 1: Difference in electronic density (red) for introducing a stacking fault into a cell of 131072 beryllium atoms (grey), simulated via machine learning.

[1]: J. A. Ellis et al., Phys. Rev. B 104, 035120, 2021
[2]: https://github.com/mala-project

1 Email: l.fiedler@hzdr.de

Physics Informed Neural Networks based Solvers for the Time- Dependent Schrödinger Equation

Karan Shah[*1], **Attila Cangi**[1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*

We demonstrate the utility of Physics Informed Neural Network based solvers for the solution of the Time-Dependent Schrödinger Equation. We study the performance and generalisability of PINN solvers on a simple quantum system. The method developed here can be potentially extended as a surrogate model for Time-Dependent Density Functional Theory, enabling the simulation of large-scale calculations of electron dynamics in matter exposed to strong electromagnetic fields, high temperatures, and pressures.

[*] Email: k.shah@hzdr.de

Fast models for warm dense matter

Timothy Callow [*1], Eli Kraisler[2], Attila Cangi[1]

[1]*Center for Advanced Systems Understanding, Helmholtz–Zentrum Dresden–Rossendorf, 02826 Görlitz, Germany*
[2]*The Hebrew University of Jerusalem, 9091401 Jerusalem, Israel*

The study of warm dense matter (WDM) is critical to our understanding of many interesting scientific and technological phenomena, in particular various astrophysical applications, and inertial confinement fusion. To develop accurate models for WDM, one has to account for the quantum behaviour of electrons (and sometimes nuclei too) across a wide range of temperatures and densities, which presents a challenge for established modelling techniques. In our poster, we introduce the concept of an average-atom model, which accounts (partially) for these quantum interactions in a computationally efficient way. We show some example applications of average-atom models, to demonstrate their usefulness in the WDM regime. We also present atoMEC: an average-atom code for matter under extreme conditions, which is open-source and written in Python.

[*] Email: t.callow@hzdr.de

Transport Properties of Matter under Extreme Conditions

Kushal Ramakrishna*[1], **Attila Cangi** [1]

[1] *Center for Advanced Systems Understanding, Helmholtz–Zentrum Dresden–Rossendorf, 02826 Görlitz, Germany*

Understanding the electronic transport properties of iron under high temperatures and pressures is essential for constraining geophysical processes. The difficulty of reliably measuring these properties under Earth-core conditions calls for sophisticated theoretical methods that can support diagnostics. We compute results of the electrical conductivity within the pressure and temperature ranges found in Earth's core from simulating microscopic Ohm's law using time-dependent density functional theory (TDDFT).
We are working on Spectral Neighbor Analysis Potential (SNAP) machine-learning potential for large-scale molecular dynamics simulations including coupling spin-lattice dynamics. The generated models can be used to simulate phenomena in iron, such as the interplay of phonon, and magnetic contributions to the thermal conductivity, or to perform high-pressure shock compression simulations.

Keywords: Density Functional Theory, Time-Dependent Density Functional Theory, Molecular Dynamics, Machine-learning Inter-atomic Potentials, Matter under Extreme Conditions

* Email: k.ramakrishna@hzdr.de

Where2Test - a digital platform for the study of COVID-19

Adam Mertel[*1], **Wildan Abdussalam**[1], **Giuseppe Barbieri**[1], **Ana Batista**[1], **Mansoor Davoodi**[1], **Kai Fan**[1], **Nishant Kumar**[1], **Abhishek Senapati**[1], **Lennart Schüler**[1,2], **Jiří Vyskočil**[1], **Weronika Schlechte-Wełnicz**[1], **Justin M. Calabrese**[1,3,4]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*
[2] *Department of Computational Hydrosystems, Helmholtz Centre for Environmental Research-UFZ, 04318 Leipzig, Germany*
[3] *Department of Ecological Modelling, Helmholtz Centre for Environmental Research-UFZ, 04318 Leipzig, Germany*
[4] *Department of Biology, University of Maryland, 20742 College Park, MD, USA*

The COVID-19 pandemic has entered a new endemic phase with the recent, highly transmissible omicron variant. To navigate this transition, the Where2Test group develops an integrated suite of models, datasets, optimization algorithms, and user-friendly webapps to help users understand and manage SARS-CoV-2 infection risk in a range of different organizational settings. On the backend, an operational data store supports applications with real-time data [2]. All of these applications are available on www.where2test.de.

- **Saxonian, Czech, and Lower Silesia dashboards** visualize the historical, actual, and predicted values of COVID-19 cases in three selected regions in Central Europe.
- **Retirement Home Testing Optmizer** helps users identify how testing strategies for long-term care facilities should change to shorten the time it takes to detect an outbreak while not overburdening care staff with testing duties.
- **COVID-19 Workplace Occupancy Optimizer** [1] calculates the optimal proportion of the employees visiting the office to infection risk to overal work productivity.
- **Saxonian wastewater dashboard** explores the spatiotemporal relations between indicator values derived from sewage systems and COVID-19 incidence, as measured by conventional testing, in the neighborhood of the wastewater plants.

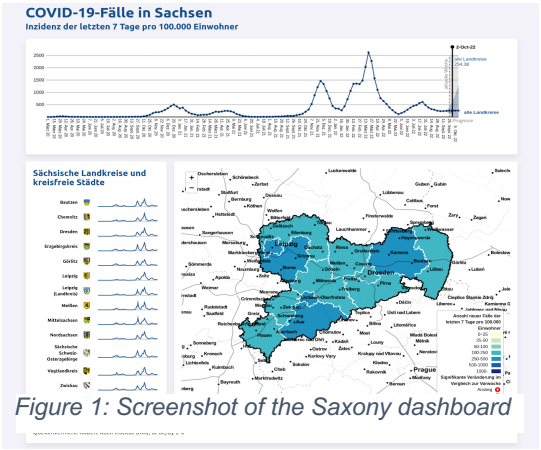


Figure 1: Screenshot of the Saxony dashboard

[1] M. Davoodi, A. Senapati, A. Mertel, W. Schlechte-Welnicz, and J. M. Calabrese, “Optimal Workplace Occupancy Strategies during the COVID-19 Pandemic,” arXiv, 2022, [Online]. Available: https://arxiv.org/pdf/2204.02062.pdf.
[2] W. Abdussalam, A. Mertel, K. Fan, L. Schüler, W. Schlechte-Wełnicz, and J. M. Calabrese, “A scalable pipeline for COVID-19: the case study of Germany, Czechia and Poland.,” 2022, [Online]. Available: https://arxiv.org/pdf/2208.12928.pdf.

* Email: a.mertel@hzdr.de

Modeling fish species diversity in river networks

Richa Tripathi[1*], **Adam Mertel**[1], **Guohuan Su**[1], **Jeffrey Kelling**[2,3], **Justin M. Calabrese**[1,4,5]

[1]*Center for Advanced Systems Understanding (CASUS), Helmholtz-Zentrum Dresden-Rossendorf,* 02826 *Görlitz, Germany*
[2] *Helmholtz-Zentrum Dresden - Rossendorf, 01314 Dresden, Germany*
[3]*Chemnitz University of Technology, 09111 Chemnitz, Germany*
[4]*Department of Biology, University of Maryland, College Park, USA*
[5]*Department of Ecological Modelling, Helmholtz Centre for Environmental Research-UFZ, 04318 Leipzig, Germany*

River basins across the world are shaped by local land topography and generally have a dendritic structure formed by convergence of river streams originating in a watershed until they end up in the main river. These river basins are also home to a plethora of aquatic lifeforms. Movement patterns of riverine biodiversity, especially fishes, are shaped by dendritic structure of river networks (see Figure 1) and habitat capacity of river basins. The ongoing river networks project at CASUS is specifically aimed at developing models to study the effects of dendritic network topology on fish biodiversity and thereby be able to predict biodiversity patterns across various river basins. Starting with an initial distribution of fish species on the river network, we explore how the biodiversity patterns, such as local species richness (LSR), in dendritic river networks evolve with time, under the assumption of species being equivalent on a *per capita* basis. Such neutral biodiversity models [1] have been able to successfully explain a suite of biodiversity indices of plant and animal species across various ecological systems [2]. In summary, the river project aims to bring together the neutral biodiversity theory and the framework of dispersal over networks to make predictions on biodiversity in riverine systems across the world. This would enable understanding the factors shaping present biodiversity and allow us to explore how climate change might affect future riverine biodiversity.

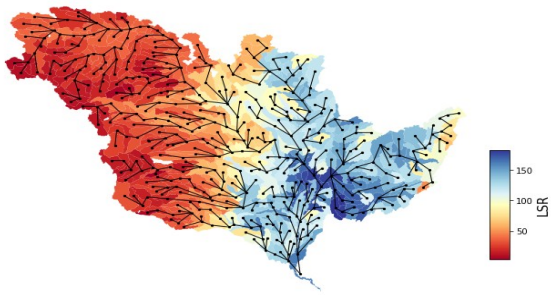


Figure 1. Mississippi-Missouri River network; the colormap showing the local species richness of river basins.

[1] Hubbell, S. P. (2011). The unified neutral theory of biodiversity and biogeography (MPB-32). In *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press.
[2] Muneepeerakul, R., Bertuzzo, E., Lynch, H. J., Fagan, W. F., Rinaldo, A., & Rodriguez-Iturbe, I. (2008). Neutral metacommunity models predict fish diversity patterns in Mississippi–Missouri basin. *Nature*, *453*(7192), 220-222.

---

*Email: r.tripathi@hzdr.de

A novel, scenario-based approach to comparing non-pharmaceutical intervention strategies across nations

Xiaoming Fu[*1], Justin M. Calabrese[1,2,3], Lennart Schüler[1,2], Sabine Attinger[2]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*
[2] *Helmholtz Centre for Environmental Research (UFZ), 04318 Leipzig, Germany*
[3] *Department of Biology, University of Maryland, College Park, USA*

Comparing the non-pharmaceutical intervention (NPI) strategies different nations employed to combat COVID-19 is a key step in preparing for future pandemics. Conventional approaches to this problem focus on identifying and ranking individual NPI effects. These efforts are complicated by vastly different political, economic, and social conditions among nations, which we refer to collectively as national framework conditions (NFCs). Furthermore, NPIs are typically applied as packages of interventions, which makes identifying their independent effects challenging. In addition, conventional approaches to studying NPI effects frequently neglect the economic and social consequences of these measures. Here, we introduce a novel, scenario-based approach to understanding NPI effects across nations. Our method couples simple epidemiological, behavioral, and economic models, and allows us to transfer NPI strategies from a reference nation to a focal nation while preserving the packaged nature of NPIs and controlling for NFCs. We conclude by considering future extensions to our framework and discussing its potential to facilitate NPI inter-comparisons worldwide.

[*] Email: x.fu@hzdr.de

Ab initio path integral Monte Carlo simulations of hydrogen snapshots at warm dense matter conditions

Maximilian Böhme [*1,2], **Zhandos A. Moldabekov**[1], **Jan Vorberger**[1]
Tobias Dornheim[1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*
[2] *Technische Universita ̈t Dresden, D-01062 Dresden, Germany*

We combine ab initio path integral Monte Carlo (PIMC) simulations with fixed ion con- figurations from density functional theory molecular dynamics (DFT-MD) simulations to solve the electronic problem for hydrogen under warm dense matter conditions [M. Bo ̈hme et al. Phys. Rev. Lett. (in print)]. The problem of path collapse due to the Coulomb at- traction is avoided by utilizing the pair approximation, which is compared against the simpler Kelbg pair-potential. We find very favourable convergence behaviour towards the former. Since we do not impose any nodal restrictions, our PIMC simulations are afflicted with the notorious fermion sign problem, which we analyse in detail. While com- putationally demanding, our results constitute an exact benchmark for other methods and approximations such as DFT. Our set-up gives us the unique capability to study im- portant properties of warm dense hydrogen such as the electronic static density response and exchange–correlation (XC) kernel without any model assumptions, which will be very valuable for a variety of applications such as the interpretation of experiments and the development of new XC functionals.

[1] D.M. Ceperley, Rev. Mod. Phys 67, 279 (1995).
[2] B. Militzer, Computer Physics Communications 204, 88–96 (2016).
[3] E.L. Pollock, Computer Physics Communications 52, 49–60 (1988).
[4] A. V. Filinov, V. O. Golubnychiy, M. Bonitz, W. Ebeling and J. W. Dufty, Phys. Rev. E 70, 046411 (2004).
[5] Tobias Dornheim, Jan Vorberger and Michael Bonitz, Phys. Rev. Lett.125, 085001 (2020).

* Email: m.boehme@hzdr.de

CASUS Professional Support Team

Jiri Vyskocil[1], **Daniel Kotik**[1], **Franz Pöschel**[1], **Giuseppe Barbieri**[1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany*

Software engineers form the CASUS Professional Support Team from diverse scientific backgrounds, including computer science, data science, and physics. With this team, the CASUS institute provides an appropriate setting for the sustainable and long-term development of scientific software, rather than being bound to the lifetime and funding of single projects.
For a digital institute focused on a cross-domain exploration of complex systems – often only possible by computational means – putting software engineering at eye level with research ensures a reliable base of high-quality software. We emphasize open-source solutions, reusability, documentation, and portability, as well as on F.A.I.R. data.
The Professional Support Team is involved in numerous in-house and cross-institutional projects:

- The scientific Python packages MALA and atoMEC are supported in the long term by code reviews, documentation generation, package management and continuous integration.
- The performance-portability framework Alpaka has given established physics simulations such as PIConGPU the chance to port to next-gen Exascale HPC systems such as ORNL Frontier. It is also used by emerging simulation and data analysis projects at CERN.
- The pandemic research platform Where2Test consists of several web applications linked to a central database containing current epidemiologic data. Predictions calculated on the HPC cluster are automatically post-processed and published online.
- Open software attracts collaboration, as the scientific I/O library openPMD-api shows, developed in collaboration between CASUS and LBNL and used by numerous scientific projects in Europe and America.
- With OPTIMA and PIONEER cloud-based platforms, the CASUS institute provides data access and analytic capabilities for researchers and pharmaceutical companies in a federated way.

The poster shows the diverse skills found within the Professional Support Team, representing the interdisciplinary nature of CASUS, and briefly introduces the projects our team is involved in.

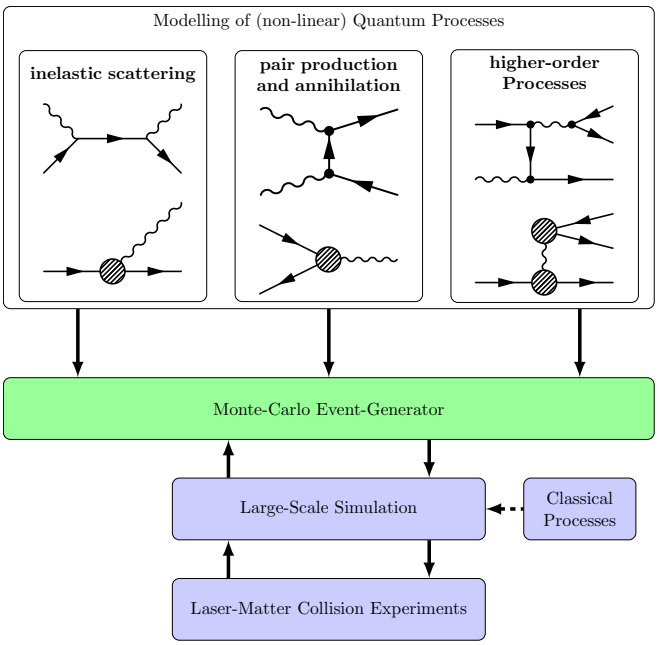
[1] Email: j.vyskocil@hzdr.de

QED.jl - Strong-field particle physics code

Klaus Steiniger[1], **Tom Jungnickel**[2], **Uwe Hernandez Acosta**[*2]

[1] Institute for Radiation Physics, *HZDR, 01328 Dresden, Germany*
[2] *Center for Advanced Systems Understanding, HZDR, 02826 Görlitz, Germany*

The collision of relativistic electron beams with highly intense, highly energetic and short-pulsed light will give deep insights into the interactions of electromagnetic fields and matter at extreme scales. Experimentally, those collisions might be examined at upcoming projects like HIBEF 2.0 at the EuropeanXFEL, SYLOS at ELI-ALPS or LCLS-II at SLAC, to name a few. The precise theoretical description of such collision experiments is very challenging and not fully covered by the currently available tools, known from particle physics. We develop the open-source software library QED.jl, which targets those gaps by providing new developments of state-of-the-art modelling tools w.r.t. strong-field physics. This includes



- **Modelling of particle physics processes:** calculation of matrix element and cross section
- **Monte-Carlo event generation:** Parallelised drawing of samples from multivariate distributions,
- **Multivariate integration:** Algorithms for highly oscillatory problems and Monte-Carlo integration for total cross sections

QED.jl is written in the Julia programming language, which opens up the usage of modern language features like just-in-time compilation, multiple-dispatch and meta-programming to attain efficiency in execution time, while the code is still easy to use and develop. Consequently, based on the computational demanding tasks given by the physics use case, necessary advances w.r.t. distributed computing are planed to be developed using Julia:

- **Task scheduling using directed acyclic graphs:** Generation of compute graphs from specific physical models, and optimisation of the evaluation of such graphs in parallel,
- **Code injection:** Extension of Julia compile workflow by injecting problem specific C++ code,
- **Hardware-agnostic parallelisation:** Kernel abstractions in Julia, e.g. supporting CPUs and GPUs from various vendors by using alpaka)

[*] Scientific supervision. Email: u.hernandez@hzdr.de

A digital twin for a nuclear waste repository in deep geologic formations

Vinzenz Brendler[*1], **Attila Cangi**[2], **Cornelius Fischer**[1], **Michael Hecht**[2], **Solveig Pospiech**[1]

[1] Institute for Resource Ecology, *Helmholtz-Zentrum Dresden-Rossendorf, 01328 Dresden, Germany*
[2] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*

A digital twin (DT) can significantly advance the site selection process for a deep geological repository for high-level radioactive waste. A respective project by the Institute of Resource Ecology, HZDR and the Center for Advanced Systems Understanding, HZDR focuses on geological, chemical, and physical aspects for the construction of such a DT. The full pathway from the radionuclide (RN) source to the ecosphere will be modeled gradually. Modeling the host and cap rocks and the adjacent damaged zone as natural compartments are great challenges, because the model complexity is accompanied by comparatively few input data.

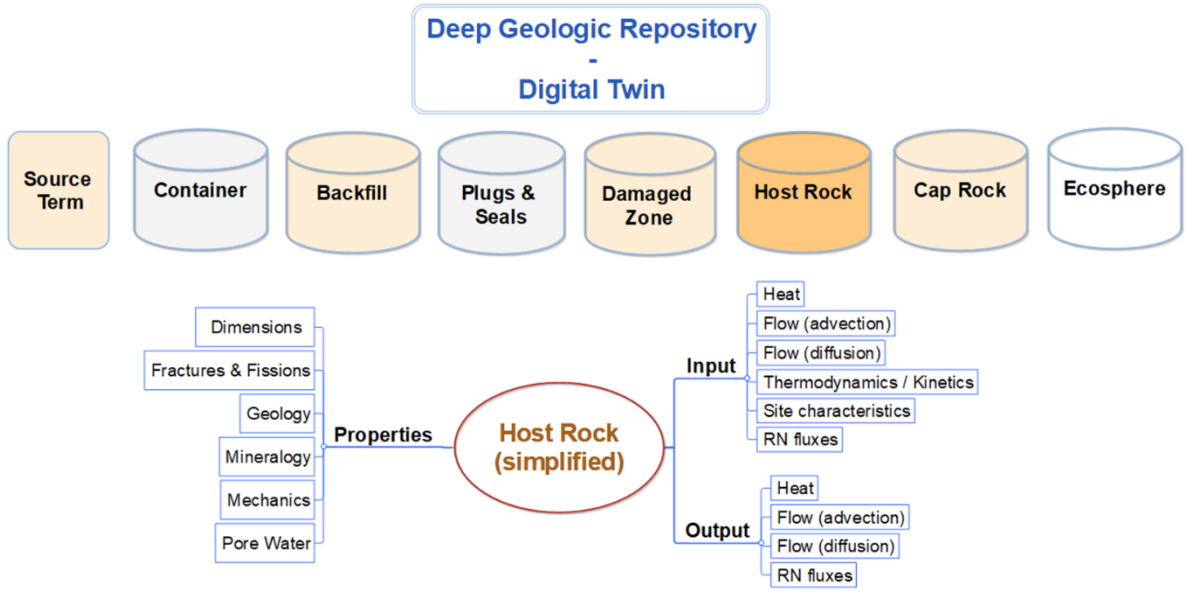


Figure 1: Essential compartments for a Digital Twin of a repository for highly radioactive waste.

Machine learning approaches will be utilized to speed up large-scale computations. A pool of machine-learning methods including Gaussian Process Regression, Kernel Ridge Regression, and Deep Neuronal Networks will particularly be used for modeling sorption processes. Training data is provided by geochemical speciation codes. A unique challenge is modeling reactive transport itself which strongly depends on the solution of differential equations. This complex problem requires a combination of Physics Informed Neural Networks with Multivariate Polynomial Interpolation. To account for data paucity and spatial heterogeneity, also geostatistics will be exploited.

* Email: v.brendler@hzdr.de

Adversarial Attacks on Aerial Vehicle Policies

Pia Hanfeld[*,1,2], **Wolfgang Hönig**[2], **Marina M.-C. Höhne**[3], **Michael Bussmann**[1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, Görlitz, Germany*
[2] *Intelligent Multi-Robot Coordination Lab, Technical University Berlin, Marchstraße 23, 10587 Berlin, Germany*
[3] *Understandable Machine Intelligence Lab, Technical University Berlin, Marchstraße 23, 10587 Berlin, Germany*

Deep Neural Networks are widely applied for solving Computer Vision tasks for Unmanned Aerial Vehicles (UAVs). For some applications, the predictions of the neural networks (NNs) directly influence the motion planning or control of the UAVs. However, the neural networks are highly prone to adversarial attacks, which has a severe negative impact on the drone's safe operation.

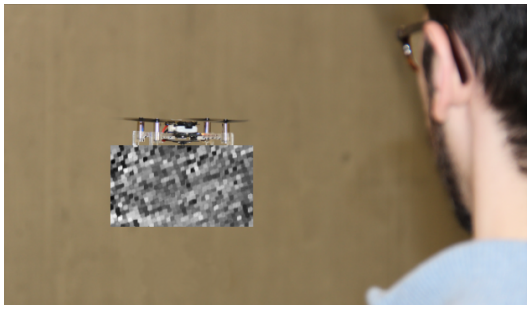


Figure 1: The attacker drone carrying a printed adversarial patch – first sketch. [1, modified]

With this work, we are planning to perform a physically realizable attack on a neural network analyzing camera images [1]. The control of the UAV is directly influenced by the predictions of this NN. The generated adversarial attacks will be printed and attached as adversarial patches to an attacker UAV. By choosing which patch to present given the current relative poses of victim and attacker, the attacker will achieve full control over the victim UAV.

[1] Palossi, D., Zimmerman, N., Burrello, A., et al. (2022). Fully Onboard AI-Powered Human-Drone Pose Estimation on Ultralow-Power Autonomous Flying Nano-UAVs. IEEE Internet of Things Journal, 9(3), 1913–1929. https://doi.org/10.1109/JIOT.2021.3091643

Email: p.hanfeld@hzdr.de; hoenig@tu-berlin.de; marina.hoehne@tu-berlin.de; m.bussmann@hzdr.de

Label-efficient Machine Learning for Diagnosing Urinary Tract Infection (UTI) in Urine Microscopy

Trina De*[1], **Artur Yakimovich**[1]

[1] *Center for Advanced Systems Understanding, Helmholtz-Zentrum Dresden-Rossendorf, 02826 Görlitz, Germany*

Urinary tract infections (UTI) belong to the most common clinically relevant bacterial infections. 1 in 3 women worldwide will have at least one UTI by 24 years of age and 40 - 50% of women will experience one UTI during their lifetime with 44% experiencing recurrences. In this project, using a clinical dataset of brightfield microscopy of patients urine with few annotated samples, we aim to develop a diagnostic phenotype quantification workflow using label-efficient machine learning (ML) approaches. There are several challenges to the clinical dataset at hand. Firstly, in the absence of specific labelling for phenotype-relevant objects in the micrographs ground truth is ambiguous. Secondly, obtaining manual annotations is laborious and requires highly-skilled annotators. Thirdly, the variation in scale and shape of a particular type of phenotype-relevant objects is challenging for the instance segmentation. To address these, first we develop a deep learning (DL) model for object detection and binary segmentation in clinical samples of patients. Since the pixel-level microscopy annotation is time-consuming and requires expert knowledge we explore label-efficient weakly or self-supervised approaches as pretext tasks to pre-train our DL model. Furthermore, to use the full extent of the optical resolution of the brightfield microscopy, as well as employ data augmentation and class balancing we use a custom generator of micrograph patches. Next, to obtain weak multi-class annotations for objects present in the micrographs we employ feature extraction with subsequent K-Means clustering. Finally, we train or fine-tune our DL model end-to-end and provide an evaluation of label-efficient techniques.

* Email: t.de@hzdr.de

Uncovering the Latent Structure of Spatiotemporal Data

Mikołaj Słupiński[1]*, **Piotr Lipiński***
* *Institute of Computer Science, University of Wrocław, Wrocław, Poland*

Many complex spatiotemporal systems can be segmented into distinct regimes with common dynamics. Discovering underlying hidden states still remains a challenging task.
This work presents the idea behind Switching Linear Dynamical Systems[1] and their variations. Those kinds of models allow for an interpretable self-supervised segmentation of spatiotemporal data and reduction of dimensionality.

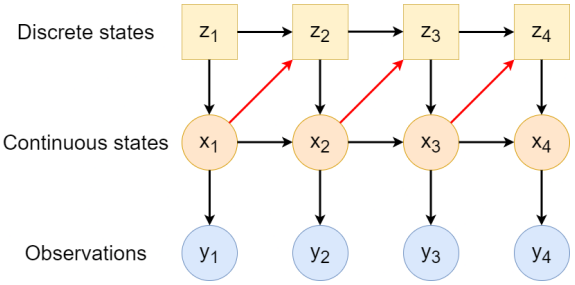


Figure 1: A graphical model of (recurrent) Switching Linear Dynamical Systems[1]. By nodes we denote random variables indexed by time, by edges we denote conditional dependences.

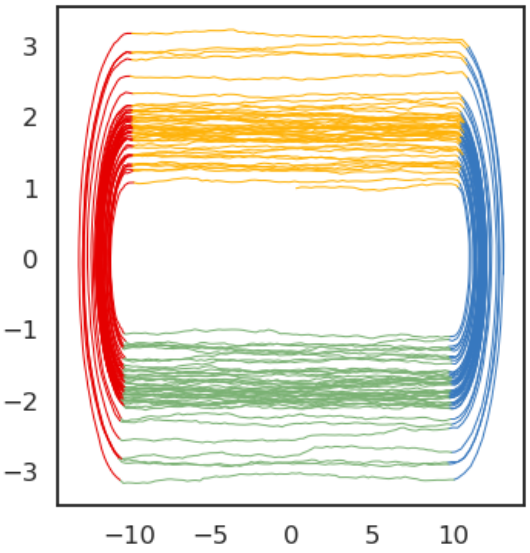


Figure 2: A simple example can be a car in a NASCAR® race. The continues state may be its position on a track. We use discrete states to influence the trajectory.

This project brings novel solutions in latent state modeling.
The main purpose of this work is to show how to create a model allowing for consistent segmentation without human input and how to train it.

The performance of the aforementioned models will be demonstrated on simple datasets.

This work is a part of a national research project on discovering hidden structure in large datasets (supported by the Polish National Science Centre (NCN) under grant OPUS-18 no. 2019/35/B/ST6/04379).

[1] Linderman, S., Johnson, M., Miller, A., Adams, R., Blei, D., Paninski, L., 2017. Bayesian Learning and Inference in Recurrent Switching Linear Dynamical Systems, in: Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. Presented at the Artificial Intelligence and Statistics, PMLR, pp. 914–922.

[1] Email: mikolaj.slupinski@cs.uni.wroc.pl

Playing Wordle with Reinforcement Learning and NLP

Adam Kaczmarek[*1]

[1] *Institute of Computer Science, University of Wrocław / ReasonField Lab*

Wordle [1] is an online game of guessing 5-letter words inspired by Mastermind. While being simply formulated it raises some questions about optimal playing strategies addressed mostly by information-theoretical [2] and tree-search [3] approaches, which while tailored for single 5-word game may not generalize well to other Wordle problem formulations like “Quordle”, “Octordle” etc. (guessing 4,8,… words simultaneously). I will present how a Reinforcement Learning approaches (Actor-Critic [4] and Policy-Gradient [5]) can be used to design a strategy for playing Wordle, as can be done for Mastermind [6]. Utilizing prior knowledge has been shown to be beneficial in RL approaches to other problems [7], [8] and Transfer-Learning is one of basis of modern approaches in Natural Language Processing. I will show that also in this case we can benefit from incorporating additional prior knowledge about language model in form of either word embeddings or character-based language model into the policy to both reduce the action space and make the exploration phase more adequate.

[1] https://www.nytimes.com/games/wordle/index.html
[2] 3Blue1Brown - Solving Wordle using Information Theory
https://www.youtube.com/watch?v=v68zYyaEmEA
[3] Jonathan Olson – Optimal Wordle Solutions
https://jonathanolson.net/experiments/optimal-wordle-solutions
[4] Vijay R. Konda John N. Tsitsiklis
Actor-Critic Algorithms
[5] Richard Sutton, David McAllester, Satinder Singh, Yishay Mansour
Policy Gradient Methods for Reinforcement Learning with Function Approximation
[6] Wei-Fu Lu; Ji-Kai Yang; Hsueh-Ting Chu
Playing Mastermind Game by Using Reinforcement Learning
[7] David L. Moreno, Carlos V. Regueiro†, Roberto Iglesias and Senen Barro
Using Prior Knowledge to Improve Reinforcement Learning in Mobile Robotics
[8] Samy Badreddine, Michael Spranger -
Injecting Prior Knowledge for Transfer Learning into Reinforcement Learning Algorithms using Logic Tensor Networks

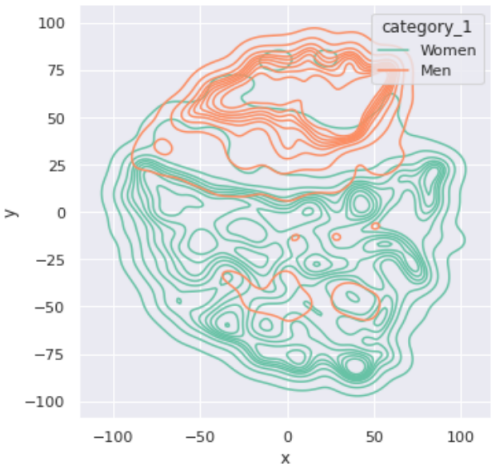
* Email: akaczmarek@cs.uni.wroc.pl / adam.kaczmarek@reasonfieldlab.com

Item Embeddings in Recommender Systems

Adrian Urbański[1*], **Maria Wyrzykowska**[*], **Paweł Rychlikowski**[*], **Piotr Lipiński**[*]

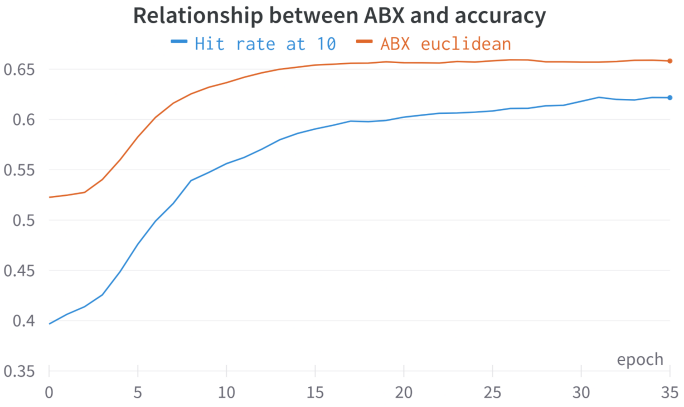
[*] *Institute of Computer Science, University of Wrocław, Wrocław, Poland*

In this research, we are exploring item embeddings in recommender systems based on deep learning approaches. In particular, we are interested in learning high quality, discriminative embeddings, as well as examining their relationship with model accuracy.



In order to generate quality item embeddings, we have used the word2vec model. In place of sentences, we provided it with sequences of items that had been interacted with by the same user. Afterwards, we calculated ABX scores on embeddings using item categories. We found out that word2vec used in such a manner generated highly discriminative embeddings.

We have also found that embeddings that effectively differentiate between categories are necessary for an accurate recommender system. We studied the relationship between the accuracy of the matrix factorization model and the ABX score of item embeddings, and observed that as the model accuracy increases, so does the ABX score.



We are currently working on applying our findings to improve predictions generated by deep learning methods, such as NeuMF [1] or TAGNN [2].

This work is a part of a national research project on discovering hidden structure in large datasets (supported by the Polish National Science Centre (NCN) under grant OPUS-18 no. 2019/35/B/ST6/04379).

[1] Xiangnan He et al., (2017). Neural Collaborative Filtering. WWW '17.
[2] Feng Yu et al., (2020). TAGNN: Target Attentive Graph Neural Networks for Session-based Recommendation. ACM SIGIR '20.

[1] Email: 308362@uwr.edu.pl

Gaussian Dense HMM

Klaudia Balcer[1*], **Paweł Rychlikowski**[*], **Piotr Lipiński**[*]

[*] *Institute of Computer Science, University of Wrocław, Wrocław, Poland*

In the research, we are working on extending dense HMMs models [1]. Our goal is to expand existing models with dense embeddings of the hidden states and discrete emissions by introducing continuous emissions and applying them to continuous problems.

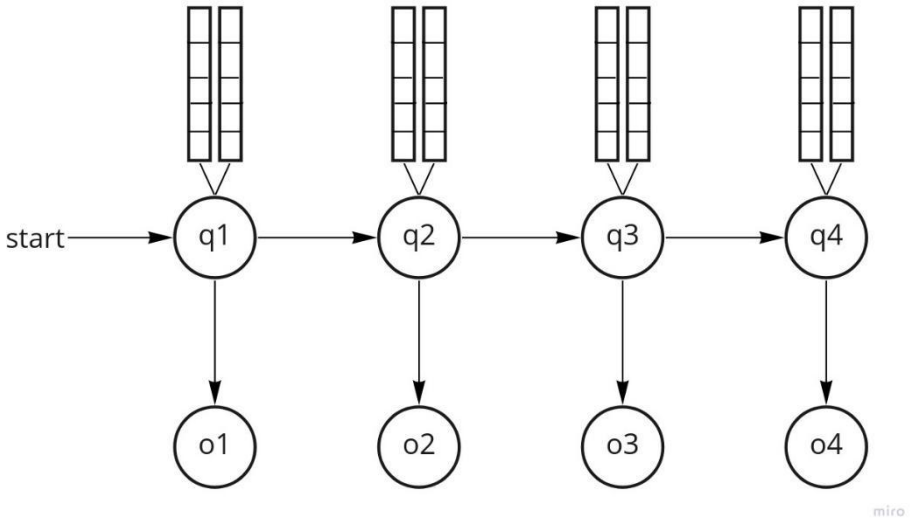


Figure 1: Model scheme.

To learn our models, we adapt a fast, constraint-free co-occurrence-based learning algorithm for continuous observation distributions. Our benchmark applications include monitoring water demand in Wroclaw's water pipeline systems and product type saturation in recommender systems.

This work is a part of a national research project on discovering hidden structure in large datasets (supported by the Polish National Science Centre (NCN) under grant OPUS-18 no. 2019/35/B/ST6/04379).

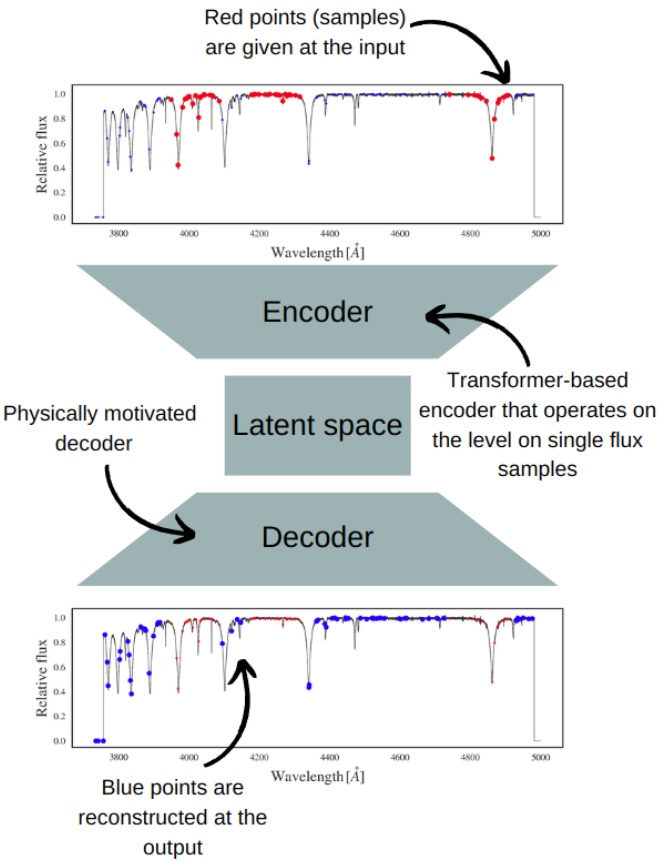
[1] J. Sicking, M.Pintz, M. Akila, T. Wirtz, "DenseHMM: Learning Hidden Markov Models by Learning Dense Representations"*, NeurIPS 2020 Workshop on Learning Meaningful Representations of Life (LMRL*), 2020

[1] Email: 300522@uwr.edu.pl

Stellar spectra classification and clustering using deep learning

Tomasz Różański[1]

[1] *Institute of Astronomy, University of Wrocław, Wrocław, Poland*



Most of our knowledge about the Universe comes from the careful analysis of light that reaches us. Spectroscopy, which is the most precise method of spectrum analysis, when applied to stars provides information on the parameters of their atmospheres, including effective temperature, acceleration, velocity fields, and their chemical composition. Stellar classification brought forth the understanding of what physical parameters are critical in shaping stellar atmospheres. It is a key element that has linked efforts related to the numerical modelling of atmospheres with observations. We present preliminary results on the method of stellar spectra classification based on large-scale unsupervised pre-training. The applied deep neural network of the autoencoder type, thanks to the use of differentiable elements of physical modelling in the decoder, allows to work with medium to high-resolution spectra, is insensitive to normalisation errors, and different radial and rotational velocity, and operates in a wide range of signal-to-noise ratio.

Email: tomasz.rozanski@uwr.edu.pl

Meshless Lattice Boltzmann Method for porous media properties calculation

Dawid Strzelczyk[*1], **Maciej Matyka**[1]

[1] *Institute of Theoretical Physics, Faculty of Physics and Astronomy, University of Wrocław, pl. Maxa Borna 9 50-204 Wrocław, Poland*

Assesing hydrodynamic parameters of porous media such as permeability or drag coefficient is an important task in various fields of science and technology, e.g. geology and heat exchangers design. At the same time complex geometries of these structures make solutions to transport equations within them impossible to be obtained analytically while numerical methods face computationally demanding process of domains discretization.

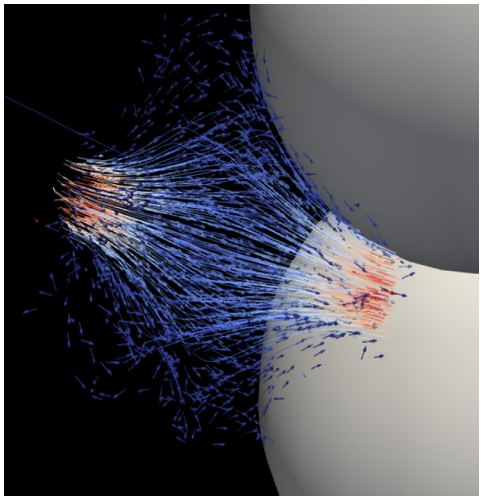


Figure 1: Streamlines and velocity vectors of flow through periodic array of spheres

An opportunity to alleviate the burden of this challenge lies in meshless methods which use discretizations comprised of irregular, non-uniform clouds of non-connected points (no need for meshing). These have been used to solve various strong form PDEs arising in fluid dynamics [1, 2].

In this poster we present an application of Meshless Lattice Boltzmann Method (ML-LBM) presented in [3] to calculate permeability and drag coefficient in a porous medium consisting of spheres arranged in a cubic lattice. We compare our results with [4] where another meshless approach – Smoothed Particle Hydrodynamics (SPH) – was used. Additionaly we briefly discuss stability of our solution in terms of varying spatial discretization of LBM model and meshless interpolation.

[1] N. Flyer, G. Wright, Journal of Computational Physics, 226(1), 2007
[2] G. Kosec, Advances in Engineering Software, 120, 2016
[3] X. Lin, J. Wu and T. Zhang, International Journal for Numerical Methods in Fluids, 91(4), 2019
[4] D. Holmes, J. WIlliams, P. Tilke, International Journal for Numerical and Analytical Methods in Geomechanics, 35(4), 2011

[*] Email: dawid.strzelczyk@uwr.edu.pl

Compute shaders in physics: sand, fluid flows, and smoothlife

Maciej Matyka
Faculty of Physics and Astronomy, University of Wrocław, pl. M. Borna 9, 50-204, Wroclaw, Poland

The rapid development of GPU (Graphics Processing Unit) technology allows to solve models in real-time simulation. CUDA and OpenCL are the most popular GPU frameworks for scientific computing. However, the low-level approach based on direct GPU programming is possible using i.e. compute shaders. See [1] where Unity3D compute shader-based simulations of autonomous agents were used.

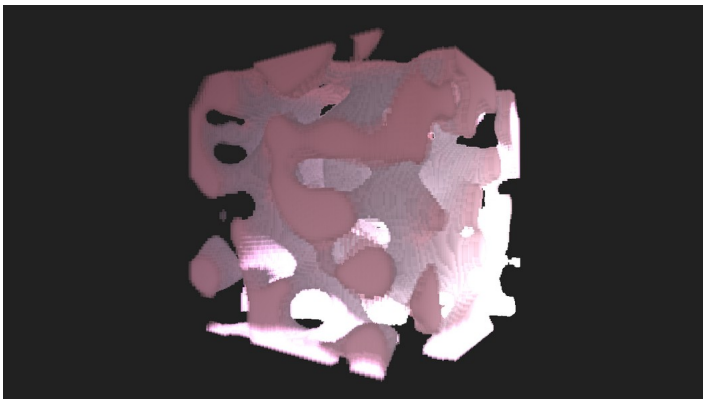


Fig. 1: 3D multiphase flow simulation visualized with ray-marching on GPU (GLSL, OpenFrameworks).

Programming directly on GPU supports real-time visualizations of results available on the graphics card. Due to our interest in real-time simulations for teaching activities and our research in porous media, we have implemented several models using compute shaders. I will present our recent work and GPU implementations of the following models:

- fluid flows through porous media with the Lattice-Boltzmann Method (LBM) [2],
- multiphase flows with LBM and ray-marching used for real-time visualization,
- sand cellular automata [3],
- smoothlife continuous variant of the game of life model [4],
- the Gray-Scott reaction-diffusion,
- solutions to classical wave equation.

Due to the specific execution model where all the data are stored on a single GPU card, this approach allowed us to get relatively high framerates in high resolutions as compared to CPU implementations.

[1] S. Bartlett and D. Louapre, Provenance of life: Chemical autonomous agents surviving through associative learning, Phys. Rev. E. 106, 034401 (2022)
[2] T. Krüger et al., The lattice Boltzmann method, Springer International Publishing 10.978-3 (2017): 4-15
[3] B. Chopard, M. Droz, Cellular Automata Modeling of Physical Systems, Cambridge University Press, 2005
[4] S. Rafler, Generalization of Conway's "Game of Life" to a continuous domain-SmoothLife, arXiv:1111.1567 (2011).

email: maciej.matyka@uwr.edu.pl

Cardinality Estimation Using Gumbel Distribution

Aleksander Łukasiewicz[*1], Przemysław Uznański[1]

[1] *Institute of Computer Science, University of Wrocław, Wrocław, Poland*

Cardinality estimation is the task of approximating the number of distinct elements in a large dataset with possibly repeating elements. LogLog and HyperLogLog (c.f. [1], [2]) are small space sketching schemes for cardinality estimation, which have both strong theoretical guarantees of performance and are highly effective in practice. This makes them a highly popular solution with many implementations in big-data systems (e.g. Algebird, Apache DataSketches, BigQuery, Presto and Redis). However, despite having simple and elegant formulation, both the analysis of LogLog and HyperLogLog are extremely involved – spanning over tens of pages of analytic combinatorics and complex function analysis.
We propose a modification to both LogLog and HyperLogLog that replaces discrete geometric distribution with the continuous Gumbel distribution. This leads to a very short, simple and elementary analysis of estimation guarantees, and smoother behavior of the estimator.

[1] Durand and Flajolet, ESA 2003
[2] Flajolet et al., Discrete Math Theor. 2007

[*] Email: aleksander.lukasiewicz@cs.uni.wroc.pl