Description of the main research directions investigated by the institute

As mentioned in the Institute's mission, the Institute main research activities concentrate around five research directions: *Computational Mathematics, Theoretical Computer Science, Machine Learning, Complex Systems, and Statistical Modelling*. Below we describe each direction in more detail.

Computational Mathematics

Computational Mathematics pursues presents a research direction studying algorithms that manipulate mathematical objects.

The field of Computational Mathematics is rapidly evolving, to a large extent due to new applications arising as mathematics and computation penetrate more and more fields of science and engineering. A recent example is the use of machine learning methods based on algorithms for large scale numerical optimization.

Such new applications have a deep influence on the field itself. For example, due to the abundance of available data and computing power the size of the problems solvable in such applications is growing immensely. This needs new, qualitatively different algorithms. Moreover, the growing complexity and heterogeneity of mathematical models makes it necessary to integrate computational methods that originally developed independently. For example, methods for solving systems of linear equations are being more and more tightly integrated into methods for solving partial differential equations.

While focusing on basic research in Computational Mathematics, the Institute actively seeks contact with such applications. The goal is to become a centre of competence in Computational Mathematics through the Institute's own excellent research that will serve as a hub for applications of Computational Mathematics in various areas.

For reaching this goal, it is necessary to have a portfolio of competences in Computational Mathematics and contact points to applications. The Institute builds on a group with a long and worldwide-acknowledged tradition in numerical linear algebra, which is *the* foundation of numerical computation. In addition, the Institute performs research in the fields of numerical optimization and verification.

In numerical optimization the research concentrates on large scale optimization and on applications, for example, in parameter estimation from data stemming from biological experiments. In the field of formal verification, we study foundations of constraint solving and of safety verification of continuous dynamical systems.

Contacts with applications are sought via various channels. For example, within the framework of Computational Mathematics we participate in the Nečas Center for mathematical modelling which is a national research platform studying applications of mathematics in continuum mechanics. Further, we also aim at applications in other fields of computer science. Especially we seek applications in fields of computer science related to the research directions of the Institute, hence exploiting synergies with the research strategy of the whole Institute.

Theoretical Computer Science

Theoretical computer science, or TCS, is a collection of interdisciplinary research areas on the borders of mathematics and computer science. Current trends in TCS are determined by a wide range of motivations ranging from purely theoretical and foundational ones stemming from the internal development of its particular areas (the notorious example being the P vs NP problem of the computational complexity) to the new ones emerging from the recent development in IT and artificial intelligence in particular. The ability to process massive data using immense computational power increases tremendously both the size and the types of the problems we can solve and also creates opportunity to tackle completely new challenges. It calls for development and study of new effective methods for processing, structuring, analysing, describing, and reasoning with the real-life information/data/ knowledge.

Within the Institute, we are focused on the more foundational aspects of the TCS and our work can be classified into three main areas: *Combinatorics, Computational Complexity,* and *Mathematical Logic*. In all these areas, we strive not only to achieve new results but to broaden the repertoire of approaches that are considered standard. In particular we pursue three main aims:

- *Combinatorics*: to study recent challenging problems in central areas of combinatorics which are crucial e.g. for our understanding of very large networks or cryptography. In the evaluated period we have mainly focused on problems from Extremal Graph Theory, Computational Geometry, and Combinatorial and Algebraic Methods in Number Theory.

- *Computational Complexity*: to deepen our knowledge of relations among complexity classes via standard and non-standard computational models considering both the usual complexity measures like time or space and new ones relevant in neurocomputing and knowledge compilation. In the evaluated period we have mainly focused on the following particular computational models: Branching Programs, Boolean formulas, and Neural Networks.

- *Mathematical Logic*: to develop and apply non-classical logics for reasoning, in both natural and artificial scenarios, with real-life information, which is often uncertain, graded, or even contradictory, and changing over time. In the evaluated period we have pursued the (abstract) algebraic study of propositional logics and are beginning to study more expressive logical formalisms, such as modal and predicate logics.

Machine Learning

Machine learning is a subfield of computer science investigating the foundations of computational systems that learn from data. It lies at the intersection of computer science and statistics. It is at the core of artificial intelligence and data science.

On one hand, the increasing availability of data has allowed machine learning systems to be trained for a large pool of applications. On the other hand, increasing computer processing power has supported the processing of huge data sets (big data problems). Consequently, machine learning systems can now outperform humans at some specific nontrivial tasks. In science, the possibility to analyse immense amount of real-world data has brought new insights in diverse areas of scientific research.

Many high-end machine learning applications are based on the intensive use of supervised learning methods, most recently applied mainly in deep neural networks, and in the structures optimized by biologically motivated methods such as genetic algorithms. Despite the indisputable success of the use of machine learning in practice, the theoretical foundations of these methods are not sufficiently explored. In particular, this concerns the paradigm of the deep neural networks. For example, a primary theoretical analysis comparing shallow and deep networks is only at its beginning and the capabilities and limits of deep networks are mainly subjects of empirical studies only. At the same time, the equally important analysis of related learning contexts needs to be exactly processed.

Answering these shortcomings, the research direction of machine learning concentrates its research effort to a number of areas: the study of the trade offs between deep network complexity and output function variability, complexity comparisons of shallow and deep networks with equivalent output, optimization of the deep network architectures design by evolutionary methods, probability modelling of the reliability of network learning, optimization of data pre-processing, sensitivity to outliers and noise, elimination of confusing inputs (classified into the class without link to this class) and characterization of tasks that can be solved by significantly simpler machine learning models.

Machine learning has become one of the key technologies of artificial intelligence, which has been in increased demand across a range of fields. The respective research cannot proceed without establishing strong theoretical foundations of these topics. The long-term tradition of the Institute in this research builds good expectations for its reasonable contribution to this field at international level.

Statistical Modelling

Statistics in its various forms, such as e.g. mathematical statistics, biostatistics, applied statistics, econometrics, psychometrics, or even bioinformatics and ‘data science’, represents one of the branches of information sciences which use data to draw inferences about reality. This is specifically achieved using the concept of designed experiments which relates to different fields of research, such as agriculture (e.g. R. A. Fischer’s experiments), biomedical sciences (providing ‘big data’ from genomic studies), medicine and pharmacology (e.g. involves statistical methodology for analysing data from clinical trials), engineering, economy (econometrics), social sciences (incl. psychology, pedagogy), and many other fields of scientific research. Observational studies, whether cross-sectional or longitudinal, offer another source of information, the latter rendering time series data. Observational studies may also involve ‘big data’ (e.g., data from satellite worldwide weather monitoring). Consequently, statistical science spans the realm of both basic and applied interdisciplinary research.

Within the corresponding research direction various topics of statistical science are being researched in the Institute. They include the following: linear and generalised linear models for independent data, mixed variants of the two classes of models for correlated data (e.g. cross-sectional and epidemiological studies), generalised additive models (GAM) allowing for flexible covariate modelling, methods for dimensionality reduction (big data), methods for analysing censored data (longitudinal, cohort studies), missing data imputation, hierarchical regression models, robust methods designed to draw correct inference in the presence of outliers, change-point problems in time series, etc. In recent years, the volume and complexity of data has increased dramatically.

Mathematical statistics is one of the building blocks that is reflected in data management and the processing and extraction of useful information from the data while controlling the level of confidence in our conclusions that are often based on large volumes of complex and noisy inputs. The ‘modelling emphasis’ indicates our interest in developing advanced methods and algorithms which take advantage from the design phase of an experiment and lead to optimised decision support in various scientific disciplines mentioned above. The department puts a key emphasis on promoting statistical approaches to solving various problems of basic and interdisciplinary research where the inspiration for developing advanced statistical methodology and algorithms comes primarily from related fields of research. Propositions arising from related scientific disciplines are reformulated as statistical hypotheses and new testing procedures are developed to answer the scientific queries.

The research in statistical modelling contributes to a further development of statistical theory and methodology, primarily in relation to the above-mentioned scientific disciplines and principal research topics. Interdisciplinary collaboration with colleagues from social sciences results in developing new methods for validation of knowledge and psychologic tests, sensitivity analysis for items and tests in relation to different target groups, and adaptive testing and mixed regression modelling. New approaches for testing structural changes in time series and panel data are under development. A part of the corresponding research efforts have been biomedical research collaboration, involving e.g. analysis of congenital defects, research of methods and algorithms for the analysis of categorical data (exact and interval estimation), and the development of advanced statistical methods for the analysis of trends in mortality and its short-term and long-term prediction. Methods to use large amounts of noisy data, often stemming from heterogeneous sources with various prior assumptions, have also been developed and refined. Attention has also been paid to the environmental data where statistical modelling can serve as a common framework for working with huge amounts of data from satellites, climate and meteorological models and various observational networks. Dynamic and/or spatial statistical modelling, of either non- or semi-parametric form, have also been at the centre of our interest. For the respective data, frequentist as well as Bayesian inference from dynamic and spatial models have been drawn, involving models identification and parameter estimation, with applications in e.g. ecology, meteorology, anthropology, medicine and technical disciplines. Dynamic and spatial inference has been applied to e.g. energy consumption modelling. The research activities in the Department of Statistical Modelling also include GAM and state-space modelling for vibrodiagnostics, as part of the Czech Academy of Sciences’ Programme ‘Strategy 21’.

Complex Systems

In general, the research of complex systems focuses on the development and application of methods for analysis and modelling of complex real-world systems. Behaviour of complex systems, typically consisting of many interacting elements, cannot be explained by a simple extrapolation of the laws describing the behaviour of a few elements. Therefore, it is important to investigate how relationships between system's parts give rise to its collective behaviour and how the system interacts and forms relationships with its environment. The study of complex systems regards collective or system-wide behaviour as the fundamental object of study. Hence, complex systems can be understood as an alternative paradigm to reductionism, which attempts to explain systems in terms of their constituent parts and individual interactions between them.

As an interdisciplinary domain, complex systems draw contributions from many different fields, such as the study of self-organization from physics, spontaneous order from the social sciences, chaos theory and graph theory from mathematics, adaptation from biology, and many others. Complex systems, as a research field, strives to solve open problems in diverse disciplines, including statistical physics, information theory, nonlinear dynamics, anthropology, computer science, Earth sciences, sociology, economics, psychology, and biology.

Within the respective research, the team of the Department of Complex Systems concentrates to the key trends corresponding to the main scientific and technological challenges:

i) the need for reliable methods for complex systems analysis,

ii) to strive for understanding the systems at multiple spatial and temporal scales,

iii) the availability of increasingly large observational/experimental datasets and the related need for employment of high-performance computing and effective algorithms,

iv) the need for integration of data driven and theory-driven approaches in complex system modelling, and

v) the ever increasing interdisciplinarity related to modelling interaction between complex systems.

Along these lines, the team focuses on high-impact research especially in selected key problem areas:

1) the development and thorough assessment of emerging complex systems data analysis methods and their utilization in societally relevant application areas,

2) the study of interactions across temporal scales in key application areas,

3) efficient computation based on a large amount of heterogeneous data leading to high computational demands. The problem includes areas of high-performance computing as well as analysis of algorithm complexity,

4) the development and optimization of frameworks integrating data and realistic theoretical priors, and

5) developing physically realistic models that bridge the gap between traditional application fields.

The main challenges in the team's research emphasize the interdisciplinary nature and usefulness in both theoretical complex systems science as well as in particular applications themselves. This makes the research both theoretical as well as close to the application field. The key methodological expertise applied includes nonlinear dynamics, information theory, network theory, stochastic processes, high-performance computing and other disciplines of computer science or related fields.

Research activity and characterisation of the main scientific results

The research activities of the team covered the three main research areas already mentioned in the description of the focus of the team: numerical linear algebra, numerical optimization, and formal verification. In the following sections, we will first discuss some of the results achieved in those areas in more detail, and then our work on applications in other areas. Each section will finish with a list of cited references.

But before going into details, we note that the rules for submitting selected publications in the first evaluation phase allowed for handing in a smaller than the maximum number of publications, in order to take into account the lower average publication counts in certain fields. Indeed, mathematics is one of the fields with the lowest average publication counts of all of science. Despite this, we decided to hand in the maximum number (in our case 21) of outputs, due to our confidence in the high quality of our publications.

Results in Numerical Linear Algebra

Since numerical linear algebra represents the central focus of the team, we will start by describing the work in this area. Numerical linear algebra forms the basis of the vast majority of numerical algorithms that - due to their ability to replace and extend physical experiments by computer simulations - have become central both for many forms of product development in industry and for many scientific areas.

In our work on numerical linear algebra, the focus was on iterative methods for solving linear systems of equations, on total least squares, and on theoretical foundations, each of which will now describe in more detail.

An important class of iterative methods are Krylov subspace methods. They belong to the most important algorithms in numerical analysis, being, for example, an item on J. Dongarra and F. Sullivan's list of 10 algorithms with the greatest influence on science and engineering. The convergence behavior of those methods has been an important topic in numerical linear algebra for many years. A long-lasting collaboration of a team member (J. Duintjer Tebbens) with a leading French researcher in numerical linear algebra (G. Meurant) significantly advanced the state of the art in this area [1, 2, 3, 4].

Another important class of iterative methods for solving linear systems of equations are matrix splitting methods which encompass, for example, the classical Jacobi and Gauss-Seidel methods. Such methods can be implemented using two iteration schemes that are mathematically equivalent, but produce different results in finite arithmetic. A team member (M. Rozložník), together with Z. Bai from the Chinese Academy of Sciences contributed a theoretical result together with mathematical experiments that settle the question of which iteration scheme is more accurate in finite arithmetic [5].

An essential part of the application of iterative methods is preconditioning, a preprocessing step for linear systems that makes them easier to solve. Several team members contributed to the design of efficient and robust preconditioning techniques [6, 7, 8, 9], collaborating with researchers such as R. Bru (Universitat Politècnica de València), and J. Scott (STFC Rutherford Appleton Laboratory and the University of Reading).

An important application of numerical linear algebra in data analysis is in methods for solving linear least squares problems. However, the class of linear total least squares problems with multiple right-hand sides have been insufficiently understood until recently. So-called core problems represent the essential information needed for solving such problems, separating it from redundancies and irrelevant information in the data. The team contributed to the generalization of the theory of such core problems from the single to the multiple right-hand side case [10, 11, 12, 13].

Team members also contributed to the theoretical foundations of numerical linear algebra, mainly to the theory of matrix factorizations [14, 15], of special matrices [16, 17, 18, 19, 20, 21], and of interval matrices [22].

References:

[1] Meurant, G., Duintjer Tebbens, Jurjen. The Role Eigenvalues Play in Forming GMRES Residual Norms with Non-Normal Matrices. *Numerical Algorithms.* 2015, 68(1), 143-165. ISSN 1017-1398. Available: doi: 10.1007/s11075-014-9891-3

[2] Duintjer Tebbens, Jurjen, Meurant, G. On the Convergence of Q-OR and Q-MR Krylov Methods for Solving Nonsymmetric Linear Systems. *Bit.* 2016, 56(1), 77-97. ISSN 0006-3835. Available: doi: 10.1007/s10543-015-0564-y

[3] Du, K., Duintjer Tebbens, Jurjen, Meurant, G. Any Admissible Harmonic Ritz Value Set is Possible for GMRES. *Electronic Transactions on Numerical Analysis.* 2017, 47(September 18), 37-56. ISSN 1068-9613. Available: doi: 10.1553/etna_vol47s37

[4] Duintjer Tebbens, Jurjen, Meurant, G. On the residual norms, the Ritz values and the harmonic Ritz values that can be generated by restarted GMRES. *Numerical Algorithms.* 2019, Online 12 December 2019. ISSN 1017-1398. Available: doi: 10.1007/s11075-019-00846-z

[5] Bai, Z.-Z., Rozložník, Miroslav. On the Numerical Behavior of Matrix Splitting Iteration Methods for Solving Linear Systems. *SIAM Journal on Numerical Analysis.* 2015, 53(4), 1716-1737. ISSN 0036-1429. Available: doi: 10.1137/140987936

[6] Kopal, J., Rozložník, Miroslav, Tůma, Miroslav. Approximate Inverse Preconditioners with Adaptive Dropping. *Advances in Engineering Software.* 2015, 84(June), 13-20. ISSN 0965-9978. Available: doi: 10.1016/j.advengsoft.2015.01.006

[7] Bru, R., Cerdán, J., Marín, J., Mas, J., Tůma, Miroslav. Preconditioners Based on the ISM Factorization. *Analele Stiintifice ale Universitatii Ovidius Constanta-Seria Matematica.* 2015, 23(3), 17-27. ISSN 1224-1784. Available: doi: 10.1515/auom-2015-0044

[8] Kopal, Jiří, Rozložník, Miroslav, Tůma, Miroslav. An Adaptive Multilevel Factorized Sparse Approximate Inverse Preconditioning. *Advances in Engineering Software.* 2017, 113(November), 19-24. ISSN 0965-9978. Available: doi: 10.1016/j.advengsoft.2016.10.005

[9] Scott, J., Tůma, Miroslav. Improving the Stability and Robustness of Incomplete Symmetric Indefinite Factorization Preconditioners. *Numerical Linear Algebra with Applications.* 2017, 24(5), e2099. ISSN 1070-5325. Available: doi: 10.1002/nla.2099

[10] Hnětynková, Iveta, Plešinger, M., Strakoš, Z. Band Generalization of the Golub-Kahan Bidiagonalization, Generalized Jacobi Matrices, and the Core Problem. *SIAM Journal on Matrix Analysis and Applications.* 2015, 36(2), 417-434. ISSN 0895-4798. Available: doi: 10.1137/140968914

[11] Hnětynková, Iveta, Plešinger, M., Sima, D.M. Solvability of the Core Problem with Multiple Right-Hand Sides in the TLS Sense. *SIAM Journal on Matrix Analysis and Applications.* 2016, 37(3), 861-876. ISSN 0895-4798. Available: doi: 10.1137/15M1028339

[12] Hnětynková, I., Plešinger, Martin, Žáková, J. Filter Factors of Truncated TLS Regularization with Multiple Observations. *Applications of Mathematics.* 2017, 62(2), 105-120. ISSN 0862-7940. Available: doi: 10.21136/AM.2017.0228-16

[13] Hnětynková, I., Plešinger, Martin, Žáková, J. Solvability Classes for Core Problems in Matrix Total Least Squares Minimization. *Applications of Mathematics.* 2019, 64(2), 103-128. ISSN 0862-7940. Available: doi: 10.21136/AM.2019.0252-18

[14] Fassbender, H., Rozložník, Miroslav. On the Conditioning of Factors in the SR Decomposition. *Linear Algebra and Its Applications.* 2016, 505(15 September), 224-244. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2016.04.038

[15] Rozložník, Miroslav, Okulicka-Dłużewska, F., Smoktunowicz, A. Cholesky-like Factorization of Symmetric Indefinite Matrices and Orthogonalization with Respect to Bilinear Forms. *SIAM Journal on Matrix Analysis and Applications.* 2015, 36(2), 727-751. ISSN 0895-4798. Available: doi: 10.1137/130947003

[16] Hnětynková, Iveta, Plešinger, M. Complex Wedge-Shaped Matrices: A Generalization of Jacobi Matrices. *Linear Algebra and Its Applications.* 2015, 487(15 December), 203-219. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2015.09.017

[17] Fiedler, Miroslav. Majorization in Euclidean Geometry and Beyond. *Linear Algebra and Its Applications.* 2015, 466(1 February), 233-240. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2014.10.015

[18] Fiedler, Miroslav. Old and New about Positive Definite Matrices. *Linear Algebra and Its Applications.* 2015, 484(1 November), 496-503. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2015.07.016

[19] Hall, F.J., Rozložník, Miroslav. G-matrices, J-orthogonal Matrices, and their Sign Patterns. *Czechoslovak Mathematical Journal.* 2016, 66(3), 653-670. ISSN 0011-4642. Available: doi: 10.1007/s10587-016-0284-8

[20] Fiedler, Miroslav. A new look at totally positive matrices. *Czechoslovak Mathematical Journal.* 2016, 66(3), 597-602. ISSN 0011-4642. Available: doi: 10.1007/s10587-016-0280-z

[21] Rohn, Jiří. Diagonally Singularizable Matrices. *Linear Algebra and Its Applications.* 2018, 555(15 October), 84-91. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2018.06.010

[22] Rohn, Jiří, Shary, S.P. Interval matrices: Regularity generates singularity. *Linear Algebra and Its Applications.* 2018, 540(1 March), 149-159. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2017.11.020

Results in Numerical Optimization

The team work in numerical optimization builds on a smaller part of the team, and hence can be described in one paragraph. Here the team concentrated on large scale optimization, reflecting the fact that numerical optimization is nowadays applied to more and more complex applications problems from various areas. The work in the area was centered around improving the efficiency of limited memory quasi-Newton methods for unconstrained optimization [23, 24, 25] and for solving non-linear systems of equations [26].

References:

[23] Vlček, Jan, Lukšan, Ladislav. A Modified Limited-Memory BNS Method for Unconstrained Minimization Based on the Conjugate Directions Idea. *Optimization Methods & Software.* 2015, 30(3), 616-633. ISSN 1055-6788. Available: doi: 10.1080/10556788.2014.955101

[24] Vlček, Jan, Lukšan, Ladislav. Properties of the block BFGS update and its application to the limited-memory block BNS method for unconstrained minimization. *Numerical Algorithms.* 2019, 80(3), 957-987. ISSN 1017-1398. Available: doi: 10.1007/s11075-018-0513-3

[25] Vlček, Jan, Lukšan, Ladislav. A Limited-memory Optimization Method Using the Infinitely Many Times Repeated BNS Update and Conjugate Directions. *Journal of Computational and Applied Mathematics.* 2019, 351(1 May), 14-28. ISSN 0377-0427. Available: doi: 10.1016/j.cam.2018.10.054

[26] Lukšan, Ladislav, Vlček, Jan. New Quasi-Newton Method for Solving Systems of Nonlinear Equations. *Applications of Mathematics.* 2017, 62(2), 121-134. ISSN 0862-7940. Available: doi: 10.21136/AM.2017.0253-16

Results in Formal Verification

Numerical algorithms form the basis of many computer simulations performed in science and industry. However, this may be hampered by problems with reliability, for example, due to imprecision introduced by numerical algorithms, and due to the necessity to select finitely many simulation scenarios to understand the infinite global behavior of the given system. Formal verification addresses such issues, with the goal of developing algorithms which provide mathematical proofs that the models of formal systems have certain desired properties.

Our work in formal verification followed two main directions during the evaluation period, the analysis of algebraic constraint solving problems under uncertainty, and the verification of dynamical systems.

In the first of those directions the team researchers answered questions such as the following: Under perturbations of which size does the solvability of a linear system of equations and inequalities stay unchanged [27]? Is it possible to check algorithmically whether a given system of non-linear equations has a robust solution [28]? How can one describe the topological structure of the solution set of a (possibly underdetermined) system of equations [29]? How can one efficiently compute the topological degree - an important tool for verifying the existence of solutions of non-linear systems of equations [30]? How can one characterize decidability of the predicate logical theory of real numbers in terms of robustness [31]?

A system of ordinary differential equations is called safe iff every solution starting in a given subset of the state space always stays in another given subset of the state space. In the work on verification of dynamical systems we developed methods for finding counter-examples to this safety property [32, 33]. Team researchers also proved that, in analogy to Lyapunov converse theorems for stability, for every robust safety verification problem for ordinary differential equations, there is a corresponding mathematical object (a so-called barrier certificate) that represents a proof of this fact [34].

References:

[27] Hladík, M., Rohn, Jiří. Radii of Solvability and Unsolvability of Linear Systems. *Linear Algebra and Its Applications.* 2016, 503(15 August), 120-134. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2016.03.028

[28] Franek, Peter, Krčál, M. Robust Satisfiability of Systems of Equations. *Journal of the ACM.* 2015, 62(4), Article 26. ISSN 0004-5411. Available: doi: 10.1145/2751524

[29] Franek, Peter, Krčál, M. On Computability and Triviality of Well Groups. *Discrete & Computational Geometry.* 2016, 56(1), 126-164. ISSN 0179-5376. Available: doi: 10.1007/s00454-016-9794-2

[30] Franek, Peter, Ratschan, Stefan. Effective Topological Degree Computation Based on Interval Arithmetic. *Mathematics of Computation.* 2015, 84(293), 1265-1290. ISSN 0025-5718. Available: doi: 10.1090/S0025-5718-2014-02877-9

[31] Franek, Peter, Ratschan, Stefan, Zgliczynski, P. Quasi-Decidability of a Fragment of the First-Order Theory of Real Numbers. *Journal of Automated Reasoning.* 2016, 57(2), 157-185. ISSN 0168-7433. Available: doi: 10.1007/s10817-015-9351-3

[32] Kuřátko, Jan. Factorization of Saddle-point Matrices in Dynamical Systems Optimization - Reusing Pivots. *Linear Algebra and Its Applications.* 2019, 566 (1 April), 61-85. ISSN 0024-3795. Available: doi: 10.1016/j.laa.2018.12.026

[33] Kuřátko, Jan, Ratschan, Stefan. Solving Reachability Problems by a Scalable Constrained Optimization Method. *Optimization and Engineering.* 2020, 21(1), 215-239. ISSN 1389-4420. Available: doi: 10.1007/s11081-019-09441-6

[34] Ratschan, Stefan. Converse Theorems for Safety and Barrier Certificates. *IEEE Transactions on Automatic Control.* 2018, 63(8), 2628-2632. ISSN 0018-9286. Available: doi: 10.1109/TAC.2018.2792325

Results in Applications

The description of our research activities can be concluded by those activities that relate our main research areas to applications both within mathematics and outside of mathematics.

A main direction here is the application of numerical linear algebra within other areas of mathematics. One of the most important contributions of computational mathematics is in giving solutions for partial differential equations which allow for the replacement of physical experiments by computer simulations in many scientific areas. The standard way of solving linear partial differential equations is to reduce them to a system of linear equations by discretization and then to solve the resulting system of linear equations. Traditionally, the assumption is made here that the system of linear equations is solved exactly, which is not the case in practice. Only recently, results have appeared that relate the error made when solving the linear equations to the discretization error. The team contributed several results in this direction [35, 36, 37].

The team also worked on applications outside of mathematics, especially in electronic structure calculations [38, 39] and in parameter estimation from biological and pharmacological experiments [40, 41, 42], collaborating with experts in the respective fields.

References:

[35] Papež, Jan, Strakoš, Z., Vohralík, M. Estimating and localizing the algebraic and total numerical errors using flux reconstructions. *Numerische Mathematik.* 2018, 138(3), 681-721. ISSN 0029-599X. Available: doi: 10.1007/s00211-017-0915-5

[36] Echeverría, C., Liesen, J., Szyld, D.B., Tichý, Petr. Convergence of the multiplicative Schwarz method for singularly perturbed convection-diffusion problems discretized on a Shishkin mesh. *Electronic Transactions on Numerical Analysis.* 2018, 48(February 22), 40-62. ISSN 1068-9613. Available: doi: 10.1553/etna_vol48s40

[37] Gergelits, Tomáš, Mardal, K.-A., Nielsen, B. F., Strakoš, Z. Laplacian Preconditioning of Elliptic PDEs: Localization of the Eigenvalues of the Discretized Operator. *SIAM Journal on Numerical Analysis.* 2019, 57(3), 1369-1394. ISSN 0036-1429. Available: doi: 10.1137/18M1212458

[38] Cimrman, R., Novák, Matyáš, Kolman, Radek, Tůma, Miroslav, Plešek, Jiří, Vackář, Jiří. Convergence study of isogeometric analysis based on Bezier extraction in electronic structure calculations. *Applied Mathematics and Computation.* 2018, 319(Feb), 138-152. ISSN 0096-3003. Available: doi: 10.1016/j.amc.2017.02.023

[39] Cimrman, R., Novák, Matyáš, Kolman, Radek, Tůma, Miroslav, Vackář, Jiří. Isogeometric analysis in electronic structure calculations. *Mathematics and Computers in Simulation.* 2018, 145(Mar), 125-135. ISSN 0378-4754. Available: doi: 10.1016/j.matcom.2016.05.011

[40] Matonoha, Ctirad, Papáček, Š. On the Connection and Equivalence of Two Methods for Solving an ill-posed Inverse Problem Based on FRAP Data. *Journal of Computational and Applied Mathematics.* 2015, 290(15 December), 598-608. ISSN 0377-0427. Available: doi: 10.1016/j.cam.2015.05.028

[41] Papáček, Š., Macdonald, B., Matonoha, Ctirad. Closed-Form Formulas vs. PDE Based Numerical Solution for the FRAP Data Processing: Theoretical and Practical Comparison. *Computers & Mathematics With Applications.* 2017, 73(8), 1673-1683. ISSN 0898-1221. Available: doi: 10.1016/j.camwa.2017.02.010

[42] Duintjer Tebbens, Jurjen, Matonoha, Ctirad, Matthios, A., Papáček, Š. On Parameter Estimation in an In Vitro Compartmental Model for Drug-Induced Enzyme Production in Pharmacotherapy. *Applications of Mathematics.* 2019, 64(2), 253-277. ISSN 0862-7940. Available: doi: 10.21136/AM.2019.0284-18

Research activity and characterisation of the main scientific results

This report is structured according to the three main research areas studied by the team (*Combinatorics, Computational Complexity, and Mathematical Logic*). Each of these three areas roughly delimits a corresponding informal research group in the team. Naturally, the research areas overlap: complexity issues of the logical deductive systems are studied within the team's research of *Mathematical Logic*, within *Computational Complexity* there is a stream of research using logical deductive systems to study Branching programs, and finally, *Computational Geometry* is one of the main areas of the team's research in *Combinatorics*.

Combinatorics: Combinatorics has become one of the main focuses of the team only recently, starting roughly from 2016. Yet, within this period it has established itself as a dynamic and diverse group with international recognition. For example, since then three of its members received a Junior grant from the Czech Science foundation, (Diana Piguet 2016-2018, Maria Saumell 2019-2021, Matas Šileikis 2020-2022) and it hosts a Standard grant as well (Diana Piguet 2019-2021). Furthermore, after the reorganization of the Institute in 2017 the team was strengthened by Štefan Porubský whose research in combinatorial and algebraic methods in number theory nicely fits into the newly created combinatorics group. The main research directions of the group are described below.

1) ***Extremal Graph Theory***: Graphs are basic mathematical structures. Their simplicity and discrete nature allows them to model a broad range of settings. Graphs are ubiquitous in theoretical computer science, contributing to the recent "digital revolution" which has radically transformed our daily lives. A lot of problems in graph theory are known or deemed to be computationally intractable (such as the notorious travelling salesman problem). For this reason, it is key to understand how different graph parameters interact. This is the goal of extremal graph theory which primarily studies the interaction between density-type parameters and the containment of a given structure, typically in very large graphs. One such example is the work of the team on the Loebl-Komlós-Sós conjecture, solving the conjecture for large dense graphs [1] and asymptotically solving it for sparse graphs [2, 3, 4, 5]. Classical methods in the area include the celebrated regularity method initiated by Szemerédi (for which Szemerédi obtained the Abel Prize in 2013) and the probabilistic method, initiated by Erdős. A nice example of the use of the probabilistic method is witnessed in the progress made on the tree-packing conjecture of Gyarfás from 1976 and the Ringel conjecture from 1963 by analysing a natural randomized algorithm for packing degenerate graphs [6]. The group also investigates asymptotic properties of graphs, contributing to the graph limit theory, which emerged in the last two decades and established fruitful links between combinatorics and analysis. In particular, paper [7] gives an alternative proof of the seminal result of [Komlos: Tiling Turan Theorems, Combinatorica, 2000] using the powerful new machinery of graph limits. Extremal Graph Theory is tightly connected to random graph theory, which studies the typical behaviour of large graphs. The main result of the group in this area is disproving a conjecture by DeMarco and Kahn on the so-called "infamous upper tail" in [8].

Researchers from the team working during the evaluated period in those (tightly connected) areas were Jan Grebík, Jan Hladký, Tereza Klimošová, Christos Pelekis, Diana Piguet, Israel Rocha, Václav Rozhoň, Maria Saumell, Matas Šileikis, and Tuan Tran. Their results appeared in top combinatorial journals such as the Journal of Combinatorial Theory Series B, the SIAM Journal on Discrete Mathematics, Random Structures & Algorithms, as well as in general mathematical journals, such as Advances in Mathematics or the Israel Journal of Mathematics.

2) ***Computational Geometry***: Computational Geometry is a research area in the intersection of Discrete Mathematics and Theoretical Computer Science. It is mainly concerned with the study of algorithmic problems that have a strong geometric component. While Computational Geometry only emerged in the second half of the 20th century, it has roots in classic geometry, one of the oldest areas of Mathematics. Nowadays it is a well-established field taught at many universities and represented at the top conferences in Theoretical Computer Science. The fact that geometry is one of the mathematical tools best suited to model real objects naturally yields a vast number of applications for research on Computational Geometry. The most prominent ones are Computer Graphics, Computer aided design and Manufacture (CAD/CAM), Pattern Recognition, Computational Morphology, VLSI Design, Artificial Vision, Geographic Information Systems, and Robotics.

Researchers from the team working during the evaluated period in Computational Geometry were Ankush Acharyya, Martin Balko, Ramesh K. Jallu, and Maria Saumell. Most of them joined the Department in 2019, thus most of their results will be published after the evaluated period. However, they already have a few remarkable results published in high-quality journals or conferences. One of their main areas of expertise are problems related to visibility. Using tools from visibility graphs, they made substantial progress on the famous potato peeling problem by giving an almost linear randomized approximation algorithm for the problem [9]. Their second main area of interest is the study of two fundamental structures in Computational Geometry, namely, Voronoi diagrams and Delaunay triangulations. They recently studied the Hamiltonicity properties of a general family of Delaunay graphs, and were able to prove that it is possible to relax the definition of Delaunay graphs with respect to a convex shape in such a way that the resulting graph always contains a Hamiltonian cycle [10].

3) ***Combinatorial Number Theory***: Number theory as one of the oldest branches of mathematics has topical and methodological touchpoints with a variety of other branches of mathematics. During the evaluated period, Štefan Porubský achieved interesting results in this area using various topological, algebraical, and analytical techniques.

Together with his coauthor he has developed a new topological technique to prove multi-layer analogons of the famous Dirichlet theorem on the infinity of prime numbers in arithmetic sequences [11]. Its application range stretches from general commutative rings satisfying certain conditions to sequences of their elements obeying properties extending the basic irreducibility property of the prime numbers. The result may help to better understand the generation of prime-number-like elements in more general algebraic structures, e.g. for various cryptological purposes.

In a series of papers (e.g. [12]) he extended an idempotent technique to describe in more detail the inner structure of some general commutative rings having roots in their classical number theory prototypes as modular rings or rings of algebraic integers.

Important applications of the theory of the uniform distribution are in the generation of the pseudorandom numbers and their applications via quasi-Monte Carlo methods. Another field of application is connected with a curious Steinhaus' three-distance theorem having consequences in combinatorics on words, or surprisingly in music or phyllotaxis. In [13] the authors found new estimates for distribution measures of a class of pseudorandom sequences generated via integral multiples of an irrational number and proved that they have new properties generalizing those of Sturmian sequences known in the combinatorics on words. They also proved that the three gap theorem is valid in one of the hitherto not observed non-trivial cases.

[1] Hladký, Jan, Piguet, Diana. Loebl-Komlós-Sós Conjecture: dense case. Journal of Combinatorial Theory. B. 2016, 116(January), 123-190. ISSN 0095-8956. Available: doi: 10.1016/j.jctb.2015.07.004

[2] Hladký, Jan, Komlós, J., Piguet, Diana, Simonovits, M., Stein, M., Szemerédi, E. The approximate Loebl-Komlós-Sós Conjecture I: The sparse decomposition. SIAM Journal on Discrete Mathematics. 2017, 31(2), 945-982. ISSN 0895-4801. Available: doi: 10.1137/140982842

[3] Hladký, Jan, Komlós, J., Piguet, Diana, Simonovits, M., Stein, M., Szemerédi, E. The approximate Loebl-Komlós-Sós Conjecture II: The rough structure of LKS graphs. SIAM Journal on Discrete Mathematics. 2017, 31(2), 983-1016. ISSN 0895-4801. Available: doi: 10.1137/140982854

[4] Hladký, J., Komlós, J., Piguet, Diana, Simonovits, M., Stein, M., Szemerédi, E. The Approximate Loebl-Komlos-Sos Conjecture III: The Finer Structure of LKS Graphs. SIAM Journal on Discrete Mathematics. 2017, 31(2), 1017-1071. ISSN 0895-4801. Available: doi: 10.1137/140982866

[5] Hladký, Jan, Komlós, J., Piguet, Diana, Simonovits, M., Stein, M., Szemerédi, E. The approximate Loebl-Komlós-Sós Conjecture IV: Embedding techniques and the proof of the main result. SIAM Journal on Discrete Mathematics. 2017, 31(2), 1072-1148. ISSN 0895-4801. Available: doi: 10.1137/140982878

[6] Allen, P., Böttcher, J., Hladký, Jan, Piguet, Diana. Packing degenerate graphs. Advances in Mathematics. 2019, 354(1 October), 106739. ISSN 0001-8708. Available: doi: 10.1016/j.aim.2019.106739

[7] Hladký, Jan, Hu, P., Piguet, Diana. Komlós's Tiling Theorem via Graphon Covers. Journal of Graph Theory. 2019, 90(1), 24-45. ISSN 0364-9024. Available: doi: 10.1002/jgt.22365

[8] Šileikis, Matas, Warnke, L. A counterexample to the DeMarco-Kahn Upper Tail Conjecture. Random Structures and Algorithms. 2019, 55(4), 775-794. ISSN 1042-9832. Available: doi: 10.1002/rsa.20859

[9] Cabello, S., Cibulka, J., Kynčl, J., Saumell, Maria, Valtr, P. Peeling Potatoes Near-optimally in Near-linear Time. Siam Journal on Computing. 2017, 46(5), 1574-1602. ISSN 0097-5397. Available: doi: 10.1137/16M1079695

[10] Bose, P., Cano, P., Saumell, Maria, Silveira, R. Hamiltonicity for Convex Shape Delaunay and Gabriel Graphs. In: Friggstad, Z., Sack, J.-R., Salavatipour, M., eds. Algorithms and Data Structures. Cham: Springer, 2019, 196-210. Lecture Notes in Computer Science, 11646. ISBN 978-3-030-24765-2. ISSN 0302-9743. Available: doi: 10.1007/978-3-030-24766-9_15

[11] Marko, F., Porubský, Štefan. Topological Aspects of Infinitude of Primes in Arithmetic Progressions. Colloquium Mathematicum. 2015, 140(2), 221-237. ISSN 0010-1354. Available: doi: 10.4064/cm140-2-5

[12] Porubský, Štefan. Semigroup Structure of Sets of Solutions to Equation X^m = X^s. Universitatis Scientiarum Budapestinensis de Rolando Eotvos Nominatae.

Annales. Sectio Computatorica. 2018, 48(1), 151-167. ISSN 0138-9491.
Available: http://ac.inf.elte.hu/Vol_048_2018/151_48.pdf

[13] Porubský, Štefan, Strauch, O. A common structure of n_k's for which n_k/alpha/bmod1/to x. Publicationes Mathematicae-Debrecen. 2015, 86(3-4), 493-502. ISSN 0033-3883.

Computational Complexity: The team members mainly focus on complexity aspects of three important specialized models of computation:

1) *Boolean formulas* (P. Savický)
2) *Neural networks* (J. Šíma, P. Savický)
3) *Branching programs* (S. Žák, J. Šíma).

Apart from time and memory requirements, other computational resources that are naturally provided by these models were investigated, employing various complexity measures including the size and depth of formulas [14, 15], buffer complexity [16], energy [17], analogicity [18, 19], periodicity [20], parameter precision, temporal coding etc.

1) ***Complexity and combinatorial properties of Boolean functions*** play a key role in computational complexity theory as well as in the design of circuits and chips for digital computers. A prominent representation is provided by CNF (conjunctive normal form) formulas. Methods for enumerating or generating models (satisfying assignments) of a CNF formula play a crucial role in several areas of computer science, especially in model checking. The paper [21] identified a non-trivial class of non-monotone formulas, for which the problem of model counting is sharp-P complete, but the enumeration can be done efficiently in the strongest sense that can be expected for a problem of this type. CNF encodings of different constraints are a key component of SAT (satisfiability) instances used to solve important problems in both theory and applications, including knowledge compilation, AI planning, and circuit and software verification. The group has focused on investigating CNF encodings with a high level of propagation strength that are used to achieve efficient responses to semantic queries concerning Boolean formulas. The paper [14], for example, provided a lower bound on the number of clauses in an encoding of the at-most-one constraint frequently used in instances for SAT solvers that is very close to a known upper bound. This research now continues within the Czech Science Foundation (CSF) grant GA19-19463S "Boolean Representation Languages Complete for Unit Propagation".

2) ***Neural networks*** (NNs) represent a prominent computational model achieving state-of-the-art results in artificial intelligence (deep learning), which is only partially theoretically justified. The group systematically develops analytical foundations of practical neurocomputing models within the CSF grant GA19-05704S "FoNeCo: Analytical Foundations of Neurocomputing"; for example, the group has characterised the computational power of NNS between integer and rational weights be establishing the analog neuron hierarchy, whose first two levels have been separated, and whose collapse to the third level has been shown [18, 19]. For the purpose of classifying NNs within the Chomsky hierarchy, a new elementary notion of a quasi-periodic number was introduced as a natural strong generalization of repeating decimals in positional numeral systems with non-integer bases [20]. The construction of quasi-periodic expansions with an infinite number of different remainders solves an important open problem for Salem bases. Another new promising idea arisen in the analysis of NNs is the concept of the simplest non-regular

deterministic context-free language which is a conceptual counterpart to (e.g. NP-) hard problems [19]. Last but not least, two optimal-size constructions of neural nets that implement a given finite automaton were proposed, one based on biologically inspired plausible robust synfire rings [22], and the other one with an energy-time trade-off [17] complemented by an exponential lower bound on the energy demands in terms of the input frequency.

3) ***Branching programs*** (BPs) constitute a non-uniform space-bounded model of computation which abstracts from model-dependent Turing-machine issues. This model, also known as a binary decision diagram (BDD), is extensively used in CAD software to synthesize circuits and in formal verification. An explicit exponential lower bound on the size of general BPs would resolve the fundamental LOGSPACE vs. PTIME problem. Along this line of research, the group has investigated restricted BPs, achieving deep results such as the polynomial-time construction of a hitting set for read-once BPs of width 3, attacking another notable unsolved LOGSPACE derandomization problem (partially supported by the CSF project GBP202/12/G061 "Center of Excellence - Institute for Theoretical Computer Science"). Recently, the paper [15] has introduced a consistent definition that describes logically an intuitive idea of what a BP knows about an input in the course of its computation, which provides a trade-off between logic and complexity and potentially opens a new very large area of BPs research.

[14] Kučera, P., Savický, Petr, Vorel, V. A Lower Bound on CNF Encodings of the At-most-one Constraint. Theoretical Computer Science. 2019, 762(March), 51-73. ISSN 0304-3975. Available: doi: 10.1016/j.tcs.2018.09.003

[15] Žák, Stanislav. A Logical Characteristic of Read-Once Branching Programs. Trier, 2019. 16 s. Electronic Colloquium on Computational Complexity, TR19-150. ISSN 1433-8092. Available: https://eccc.weizmann.ac.il/report/2019/150/

[16] Šíma, Jiří, Žák, Stanislav. On Tight Separation for Blum Measures Applied to Turing Machine Buffer Complexity. Fundamenta Informaticae. 2017, 152(4), 397-409. ISSN 0169-2968. Available: doi: 10.3233/FI-2017-1526

[17] Šíma, Jiří. Energy-Time Tradeoff in Recurrent Neural Nets. In: Koprinkova-Hristova, P., Mladenov, V., Kasabov, N.K., eds. Artificial Neural Networks. Methods and Applications in Bio-/Neuroinformatics. Cham: Springer, 2015, s. 51-62. Springer Series in Bio-/Neuroinformatics, 4. ISBN 978-3-319-09902-6. ISSN 2193-9349. Available: doi: 10.1007/978-3-319-09903-3_3

[18] Šíma, Jiří. Subrecursive Neural Networks. Neural Networks. 2019, 116(August), 208-223. ISSN 0893-6080. Available: doi: 10.1016/j.neunet.2019.04.019

[19] Šíma, Jiří, Plátek, Martin. One Analog Neuron Cannot Recognize Deterministic Context-Free Languages. In: Gedeon, T., Wong, K.-W., Lee, M., eds. Neural Information Processing. Proceedings, Part III. Heidelberg: Springer, 2019, 77-89. Lecture Notes on Computer Science, 11955. ISBN 978-3-030-36717-6. Available: doi: 10.1007/978-3-030-36718-3_7

[20] Šíma, Jiří, Savický, Petr. Quasi-Periodic beta-Expansions and Cut Languages. Theoretical Computer Science. 2018, 720(11 April), 1-23. ISSN 0304-3975. Available: doi: 10.1016/j.tcs.2018.02.028

[21] Savický, Petr, Kučera, P. Generating Models of a Matched Formula with a Polynomial Delay. Journal of Artificial Intelligence Research. 2016, 56(6), 379-402. ISSN 1076-9757. Available: doi: 10.1613/jair.4989

[22] Cabessa, Jérémie, Šíma, Jiří. Robust Optimal-Size Implementation of Finite State Automata with Synfire Ring-Based Neural Networks. In: Tetko, I. V., Kůrková, V., Karpov, P., Theis, F., eds. Artificial Neural Networks and Machine Learning – ICANN 2019: Theoretical Neural Computation. Proceedings, Part I. Cham: Springer, 2019, 806-818. Lecture Notes in Computer Science, 11727. ISSN 0302-9743. ISBN 978-3-030-30486-7. Available: doi: 10.1007/978-3-030-30487-4_62

Mathematical logic: Logic claims as its own an area on the crossroads of mathematics, computer science and philosophy, with applications in many other areas of human endeavour such as the social sciences, law, or linguistics. The group studies logics in their plurality and richness, with the primary research motivation, methodology, and emphasis succinctly described as a ***mathematical investigation of formal systems of non-classical logics***. Some members of the group may complement this primary motivation by two additional ones: a *study and application of logics as formal tools for various computer science problems* and a *philosophical development of logic as a science of reasoning*.

The origins of the logic group can be traced to the study of Mathematical Fuzzy Logic, a research discipline established by prof. Petr Hájek, who also founded the group; the study provided logical foundations for reasoning with vague/graded notions, thus contributing to all three motivations mentioned above. Building on this experience, research during the evaluated period has shifted and expanded. The area of interest has grown from fuzzy logics to richer classes of non-classical logics (in its limit, non-classical logics are studied as abstract mathematical objects). Computational properties of various logical systems have been studied, aimed at potential computer science applications. Moreover, extending the study to modal and predicate formalisms acknowledges the indispensability of more expressive formal tools for both computer science applications and realistic models of reasoning.

The next table contains a structured list of the members of the group and depicts its development during the evaluated period:

	2015	2016	2017	2018	2019
Tenured	Cintula Haniková Horčík	Cintula Haniková Horčík	Cintula Haniková Horčík	Cintula Haniková	Cintula Haniková
Tenure track				Bílková	Bílková
Postdoc	Suzuki Sarkoci (OCT) Petrík	Suzuki (JUL) Sarkoci Petrík Vidal (MAY) Sedlár (JUN) Moraschini (JUL)	Bonzio (APR-NOV) Petrík Vidal Sedlár Moraschini	Baldi (APR-DEC) Reggio (OCT) Tedder (SEP) Vidal Sedlár Moraschini	Reggio (JUL) Tedder Vidal Sedlár Moraschini
Part time project worker	Běhounek (MAY) Bílková Majer	Bílková Majer	Bílková Majer	Punčochář Majer	Punčochář Majer
PhD. students	Dostál Chvalovský Přenosil Lávička (OCT)	Chvalovský (OCT) Přenosil Lávička	Přenosil	Přenosil (AUG)	

Research during the evaluated period roughly followed four main lines, listed below together with the names of the team members who have investigated them:

1) *Propositional substructural logics*: Petr Cintula, Zuzana Haniková, Rostislav Horčík, Libor Běhounek, Stefano Bonzio, Karel Chvalovský, Ondrej Majer, Tommaso Moraschini, Milan Petrík, Adam Přenosil, Luca Reggio, Peter Sarkoci, Igor Sedlár, Andrew Tedder, Amanda Vidal
2) *Abstract algebraic logic*: Petr Cintula, Tomáš Lávička, Tommaso Moraschini, Adam Přenosil
3) *Modal many-valued logics*: Marta Bílková, Petr Cintula, Matěj Dostál, Vít Punčochář, Igor Sedlár, Tomoiuki Suzuki, Amanda Vidal, Andrew Tedder
4) *First-order non-classical logics*: Petr Cintula, Rostislav Horčík, Libor Běhounek, Amanda Vidal

1) ***Propositional substructural logics*** provide a unified framework for a multitude of logics, such as superintuitionistic, relevant, or many-valued logics, and can be obtained by adding various combinations of *structural rules* (associativity, contraction, exchange, weakening) to a sequent calculus for the nonassociative Full Lambek calculus (originating in categorial grammars), and possibly extending these basic substructural logics further.
   - *Computational properties of substructural logics*. Solutions of two long-standing open problems complete a mosaic of decidability results for basic substructural logics: paper [23] shows undecidability of the consequence relation of nonassociative Full Lambek calculus (by finding an encoding of the halting problem for 2-tag systems) and paper [24] shows that theoremhood in Full Lambek calculus with contraction is undecidable (combining methods from universal algebra, regular languages, and string rewriting theory). Paper [25] studies computational aspects of various satisfiability problems in algebraic semantics for the Full Lambek logic with exchange and weakening.
   - *Paraconsistent, relevant, and fuzzy propositional logics*. The focus was on the study of completeness properties, definability, and proof theory of these logics. For example the paper [26] presents a systematic Gentzen-style proof theory for the study of extensions of the famous Belnap-Dunn paraconsistent logic. Papers [27] and [28] exemplify work in fuzzy logics that expand the language of Full Lambek calculus with additional connectives to improve expressivity on propositional level.
   - *Ordered algebraic structures*. Propositional logics are strongly linked to their algebraic semantics, whose examination leads to research addressing problems and using tools typical of (universal) algebra. Members of the group published numerous papers studying such structures. E.g., paper [29] characterizes the subdirectly irreducible De Morgan monoids and proves that there are just four minimal varieties of these algebraic structures. Paper [30] shows that the universal theory of square-increasing idempotent semirings is decidable. Paper [31] uses the concept of residuated frames to provide a unified densification method for chains in the class of Full Lambek algebras.
2) ***Abstract algebraic logic*** (AAL) studies logical systems by considering the ways in which they may be associated with their algebraic counterparts. During the evaluated period the group members have published over 20 journal papers on various problems in AAL; a few particular examples are described below:
   - The paper [32] is the culmination of a series of papers studying an abstract hierarchy of non-classical logics. Developing and using a powerful framework,

building on results scattered in the literature, the paper provides strong characterization theorems for various completeness properties which link two fundamental presentations of any logic: syntax (given by proof systems) and semantics (given in terms of truth and consequence).

- Category theory studies bridges between different areas of mathematics; these bridges are typically given in terms of "adjunctions", i.e. well-behaved pairs of transformations between formalizations of distinct areas of mathematics. Using methods of universal algebra, paper [33] gives a combinatorial description of adjunctions in the realm of equational algebra and uses this result to illuminate an unexpected correspondence between adjunctions and syntactic translations between logical systems.
- Tarski's notion of "consequence relation" is the cornerstone of modern mathematical logic. However, the recent advent of various "consequence like" relations in logic and computer science (e.g. resource-aware ones where the number occurrences of premises matters) calls for a more general approach. The paper [34] proposes such a framework, relates it to existing approaches, and demonstrates its power and utility with a plethora of examples.
- The Leibniz hierarchy is a fundamental taxonomy of deductive systems in terms of the behavior of their logical equivalence relations. The problem of classifying logics within this hierarchy is shown to be undecidable for syntactically presented logics [35] and decidable but EXPTIME-complete for logics presented by a finite set of finite models [36].

3) ***Modal many-valued logics***: The combination of modal and many-valued logics is gaining more and more attention from both communities as they combine tools for representing vagueness or handling inconsistency employed by many-valued logic with tools offered by modal logics for representing actions, knowledge, belief, obligations, etc. During the evaluated period the team members have published over 20 journal papers on various modal logics; a few particular examples are described below:
   - Paper [37] provides a computer tool for deciding satisfiability, theoremhood and consequence in many-valued modal logics. It designs, implements, and exhaustively tests this tool. Its design is based on the classical Satisfiability Modulo Theories solver z3 and uses some recent advanced theoretical results on these logics to dramatically improve its computational behaviour.
   - Paper [38] extends earlier particular results on expressivity of standard modal languages over relational structures with many-valued valuations. It is written from a more general coalgebraic perspective and proves a strong characterization theorem, crucial to fully understand and explore the limits of expressivity of abstract modal languages. The results bridge three communities (that of many-valued, modal, and coalgebraic logics).
   - Paper [39] is the first study of the combination of dynamic logics used to formalize reasoning about actions (with applications e.g. in computer science for program verification) and paraconsistent logics which are used for non-trivial treatment of inconsistent information.
4) ***Predicate non-classical logics***: It is clear that for certain applications (in mathematics, computer science, or philosophy) the full expressive power of the first- (or even higher-) order predicate logics is needed. During the evaluated

period the team members have published over 10 journal papers in this area; particular examples are described below:

- Paper [40] generalized a foundational result of mathematical logic, Gödel´s completeness theorem, to a wide class of non-classical first-order logics. Its approach is based on non-trivial modifications of Henkin's proof of the classical result and, since almost no properties of the underlying propositional logics are assumed, it not only encompasses most non-classical logics studied in the literature but also contributes to a finer understanding of the classical proof.
- Herbrand and Skolem theorems for classical logic are interesting from both theoretical and applied perspectives (e.g. they are instrumental for automated deduction). The paper [41] presents a novel framework for studying various forms of these theorems in non-classical logics. The framework is used to prove several strong general characterization theorems which yield not only a plethora of new results but also new elegant proofs of a majority of known results scattered in the literature.
- Paper [42] is an example of application-driven research. The theory of constraint satisfaction problems has recently been generalized to a valued version (which turns it into an optimization problem). The paper proposed a valued version of Geiger's result that allows one to characterize the pp-definable relations over valued structures in terms of certain mappings of the structure in itself. This opens the way to a new approach to VCSP, contemplating a wider class of problems and a clearer methodology of many-valued predicate logics to study their behavior and computational complexity.

[23] Chvalovský, Karel. Undecidability of Consequence Relation in Full Non-associative Lambek Calculus. Journal of Symbolic Logic. 2015, 80(2), 567-586. ISSN 0022-4812. Available: doi: 10.1017/jsl.2014.39

[24] Chvalovský, Karel, Horčík, Rostislav. Full Lambek Calculus with Contraction is Undecidable. Journal of Symbolic Logic. 2016, 81(2), 524-540. ISSN 0022-4812. Available: doi: 10.1017/jsl.2015.18

[25] Haniková, Zuzana, Savický, Petr. Term Satisfiability in FLew-Algebras. Theoretical Computer Science. 2016, 631(6 June), 1-15. ISSN 0304-3975. Available: doi: 10.1016/j.tcs.2016.03.009

[26] Přenosil, Adam. Cut Elimination, Identity Elimination, and Interpolation in Super-Belnap Logics. Studia Logica. 2017, 105(6), 1255-1289. ISSN 0039-3215. Available: doi: 10.1007/s11225-017-9746-8

[27] Haniková, Zuzana. Implicit Definability of Truth Constants in Łukasiewicz Logic. Soft Computing. 2019, 23(7), 2279-2287. ISSN 1432-7643. Available: doi: 10.1007/s00500-018-3461-x

[28] Cintula, Petr. A Note on Axiomatizations of Pavelka-style Complete Fuzzy Logics. Fuzzy Sets and Systems. 2016, 292(1 June), 160-174. ISSN 0165-0114. Available: doi: 10.1016/j.fss.2014.11.021

[29] Moraschini, Tommaso, Raftery, J.G., Wannenburg, J. J. Varieties of De Morgan Monoids: Minimality and Irreducible Algebras. Journal of Pure and Applied Algebra. 2019, 223(7), 2780-2803. ISSN 0022-4049. Available: doi: 10.1016/j.jpaa.2018.09.015

[30] Horčík, Rostislav. On Square-Increasing Ordered Monoids and Idempotent Semirings. Semigroup Forum. 2017, 94(2), 297-313. ISSN 0037-1912. Available: doi: 10.1007/s00233-017-9853-x

[31] Baldi, Paolo, Terui, K. Densification of FL Chains via Residuated Frames. Algebra Universalis. 2016, 75(2), 169-195. ISSN 0002-5240. Available: doi: 10.1007/s00012-016-0372-5

[32] Cintula, Petr, Noguera, Carles. Implicational (semilinear) logics III: completeness properties. Archive for Mathematical Logic. 2018, 57(3-4), 391-420. ISSN 0933-5846. Available: doi: 10.1007/s00153-017-0577-0

[33] Moraschini, Tommaso. A Logical and Algebraic Characterization of Adjunctions between Generalized Quasi-Varieties. Journal of Symbolic Logic. 2018, 83(3), 899-919. ISSN 0022-4812. Available: doi: 10.1017/jsl.2018.47

[34] Cintula, Petr, Gil-Férez, J., Moraschini, Tommaso, Paoli, F. An Abstract Approach to Consequence Relations. Review of Symbolic Logic. 2019, 12(2), 331-371. ISSN 1755-0203. Available: doi: 10.1017/S175502031900008X

[35] Moraschini, Tommaso. A Computational Glimpse at the Leibniz and Frege Hierarchies. Annals of Pure and Applied Logic. 2018, 169(1), 1-20. ISSN 0168-0072. Available: doi: 10.1016/j.apal.2017.07.003

[36] Moraschini, Tommaso. On the complexity of the Leibniz hierarchy. Annals of Pure and Applied Logic. 2019, 170(7), 805-824. ISSN 0168-0072. Available: doi: 10.1016/j.apal.2019.02.003

[37] Vidal, Amanda. MNiBLoS: A SMT-based Solver for Continuous t-norm Based Logics and Some of their Modal Expansions. Information Sciences. 2016, 372(1 December), 709-730. ISSN 0020-0255. Available: doi: 10.1016/j.ins.2016.08.072

[38] Bílková, Marta, Dostál, Matěj. Expressivity of Many-Valued Modal Logics, Coalgebraically. In: Väänänen, J., Hirvonen, A., de Queiroz, R., eds. Logic, Language, Information, and Computation. Berlin: Springer, 2016, 109-124. Lecture Notes in Computer Science, 9803. ISBN 978-3-662-52920-1. ISSN 0302-9743. Available: doi: 10.1007/978-3-662-52921-8_8

[39] Sedlár, Igor. Propositional Dynamic Logic with Belnapian Truth Values. In: Beklemishev, L., Demri, S., Máté, A., eds. Advances in Modal Logic. Vol. 11. London: College Publications, 2016, s. 503-519. ISBN 978-1-84890-201-5. Available: https://www.collegepublications.co.uk/aiml/?00008

[40] Cintula, Petr, Noguera, Carles. A Henkin-Style Proof of Completeness for First-Order Algebraizable Logics. Journal of Symbolic Logic. 2015, 80(1), 341-358. ISSN 0022-4812. Available: doi: 10.1017/jsl.2014.19

[41] Cintula, Petr, Diaconescu, D., Metcalfe, G. Skolemization and Herbrand theorems for lattice-valued logics. Theoretical Computer Science. 2019, 768(10 May), 54-75. ISSN 0304-3975. Available: doi: 10.1016/j.tcs.2019.02.007

[42] Horčík, Rostislav, Moraschini, Tommaso, Vidal, Amanda. An Algebraic Approach to Valued Constraint Satisfaction. In: Goranko, V., Dam, M., eds. 26th EACSL Annual Conference on Computer Science Logic. Dagstuhl: Leibniz-Zentrum fuer Informatik, 2017. Leibniz International Proceedings in Informatics, 82. ISBN 978-3-95977-045-3. ISSN 1868-8969. Available: http://drops.dagstuhl.de/opus/volltexte/2017/7676/

Research activity and characterisation of the main scientific results

The Department of Machine Learning, which coincides with the evaluated team, was established in the middle of the evaluated period (July 2017). In the first half of the reporting period, the team members were spread across other departments. Despite this fact, the professional focus of the team members was stable throughout the evaluated period and in line with the professional focus of the newly constituted Department of Machine Learning. During the period under review, team members achieved the following results in the research areas, which were mentioned above and will be discussed in more details below.

Focus 1. Theoretical properties of machine learning methods, especially approximation potential and complexity of layered neural networks

The choice of dictionaries of computational units suitable for efficient computation of binary classification tasks has been investigated. To deal with exponentially growing sets of tasks with increasingly large domains, a probabilistic model has been introduced. The relevance of tasks for a given application area has been modeled by a probability distribution on the set of all binary classifiers on a given domain. Probabilistic estimates have been derived for approximate measures of network sparsity formalized in terms of suitable norms [1].

The limitations of capabilities of shallow networks to efficiently compute real-valued functions on finite domains have been investigated. The efficiency in terms of network sparsity and its approximate measures has been studied. It has been shown that when a dictionary of computational units is not sufficiently large, computation of almost any uniformly randomly chosen function either represents a well-conditioned task performed by a large network or an ill-conditioned task performed by a network of a moderate size. The probabilistic results are complemented by a concrete example of a class of functions which cannot be efficiently computed by shallow perceptron networks. The class is constructed using pseudo-noise sequences which have many features of random sequences but can be generated using special polynomials [2].

A concrete construction of a class of functions which cannot be computed by perceptron networks with considerably smaller numbers of units and smaller output weights than the sizes of domains of the functions has been presented. A subset of this class has been described whose elements can be computed by two hidden-layer perceptron networks with the numbers of units depending on a logarithm of the size of the domain linearly [3].

Lower bounds on errors in approximation of binary-valued functions by shallow perceptron networks have been derived. It has been proven that unless the number of network units is sufficiently large (larger than any polynomial of the logarithm of the size of the domain) a good approximation cannot be achieved for almost any uniformly randomly chosen function. The results are obtained by combining probabilistic Chernoff–Hoeffding bounds with estimates of sizes of sets of functions computable by shallow networks with increasing numbers of units [4].

Model complexities of shallow networks computing highly-varying functions have been investigated in terms of variational norms tailored to dictionaries of computational units. Characterization of distributions of these norms are derived for networks with popular computational units such as Gaussian radial units and perceptrons [5].

[1] Kůrková, Věra, Sanguineti, M. Classification by Sparse Neural Networks. IEEE Transactions on Neural Networks and Learning Systems. 2019, 30(9), 2746-2754. ISSN 2162-237X. Available: doi: 10.1109/TNNLS.2018.2888517

[2] Kůrková, Věra. Limitations of Shallow Networks Representing Finite Mappings. Neural Computing & Applications. 2019, 31(6), 1783-1792. ISSN 0941-0643. Available: doi: 10.1007/s00521-018-3680-1

[3] Kůrková, Věra. Constructive Lower Bounds on Model Complexity of Shallow Perceptron Networks. Neural Computing & Applications. 2018, 29(7), 305-315. ISSN 0941-0643. Available: doi: 10.1007/s00521-017-2965-0

[4] Kůrková, Věra, Sanguineti, M. Probabilistic Lower Bounds for Approximation by Shallow Perceptron Networks. Neural Networks. 2017, 91(July), 34-41. ISSN 0893-6080. Available: doi: 10.1016/j.neunet.2017.04.003

[5] Kůrková, Věra, Sanguineti, M. Model Complexities of Shallow Networks Representing Highly Varying Functions. Neurocomputing. 2016, 171(1 January), 598-604. ISSN 0925-2312. Available: doi: 10.1016/j.neucom.2015.07.014

Focus 2. Alternative computational models and knowledge generating systems

The field of artificial intelligence deals with a number of alternative computational models extending the set of standard computational models. For instance, to be able to theoretically handle more general concepts and tasks like general knowledge, knowledge acquisition and generation, mechanisms of new knowledge emergence, cognitive computation, etc., it is often advantageous to develop new computational and information models incorporating these concepts. The respective research can even lead to new foundations of computation not based on machine models. On the other hand, in computability theory there are still issues related to deficiencies of the classical computational models to capture classical complexity classes (such as the complements of recursively enumerable languages) in a natural way. Such investigations extend general understanding of information processing, contribute to the development of the respective theory and can find their use in future applications of computing. This is why within the department we also focus to the design of alternative models of information processing aimed at new ways of exploiting of knowledge and information.

For example, question-answering systems like Watson beat humans when it comes to processing speed and memory. But what happens if we compensate for this? What are the fundamental differences in power between human and artificial agents in question answering? This issue has been explored by defining new computational models for both agents and comparing their computational efficiency in interactive sessions [6]. It was shown that cognitive automata, modeling cognitive agents, and QA-machines modeling Watson-like machines have exactly the same potential in realizing question-answering sessions, provided the resource bounds in one model are sufficient to match the abilities of the other.

Another computational model, finite state machines with feedback, presents a novel machine model when considered under the scenario of cognitive computations. The model is designed in the spirit of automata theory [7] and presents a mix of Alan Turing's finite state machines and Norbert Wiener's machines with feedback. For the purposes of the model, a team member has defined what is called minimal machine consciousness and machine qualia. The design of the model is led by natural engineering requirements. Its properties are justified by the latest findings in neuroscience and by ideas from the classical literature of the philosophy of mind. For the model a test distinguishing minimally conscious machines from unconscious ones ("so-called philosophical zombies") on a given cognitive task is proposed. The modelling supports the claim that consciousness is a computational phenomenon that is not just a matter of suitable software but also requires a dedicated architecture.

The notion of computation is well understood, and well formalized, in the classical context of digital information processing. However, the paradigm of computation is increasingly used in the characterization of processes in sciences like physics and biology as well. This seems to recognize computation more broadly as an 'elementary mechanism of nature', not limited to the digital domain. What is computation, if it is observed in this broader sense? Without computers as the single defining mechanism, classical abstractions like Turing machines and rewrite systems seem to be unsuited to capture its newer meanings. Following an epistemic approach, a new framework for understanding computation and, as a consequence, for understanding computational processes, based on observed trajectories (curves) of computational activity in suitable metric spaces has been developed [8]. With computational processes rather than all sorts of different models of computation as the core abstraction, one can model and characterize many aspects of computation as a mathematical object, from composition to computational and structural complexity. Discrete computations appear as projections of continuous ones, clarifying complex issue from Turing's 1948 report. In [8], both the philosophy and a first outline of the implied, machine-independent, general mathematical theory of computation, is presented.

The effort has also been addressed to resolve some old problems, namely to design a 'natural' machine model for accepting the complements of recursively enumerable (RE) languages. The new model is based on non-deterministic Turing machines with 'one-sided' advice [9]. It has been proved that these machines precisely accept the co-RE languages, without restriction on the advice functions that are used. Another problem which attracted team attention is an observation that the RE languages form an infinite proper hierarchy $RE1 \subset RE2 \subset \cdots$ based on the size of the Turing machines that accept them. The fundamental position of the finite languages and their complements in the hierarchy have been proved [10]. As a new technique in this domain team members have employed various results for Turing machines 'with advice'. They have shown that optimizing the (length of the) advice in a Turing machine can be an effective intermediate step in bounding the state complexity of finite languages.

[6] van Leeuwen, J., Wiedermann, Jiří. Question Answering by Humans and Machines: A Complexity-theoretic View. Theoretical Computer Science. 2019, 777(19 July), 464-473. ISSN 0304-3975. Available: doi: 10.1016/j.tcs.2018.08.012

[7] Wiedermann, Jiří, van Leeuwen, J. Finite State Machines with Feedback: An Architecture Supporting Minimal Machine Consciousness. In: Manea, F., Martin, B., Paulusma, D., Primiero, G., eds. Computing with Foresight and Industry. Cham: Springer, 2019, 286-297. Lecture Notes in Computer Science, 11558. ISBN 978-3-030-22995-5. ISSN 0302-9743. Available: doi: 10.1007/978-3-030-22996-2_25

[8] van Leeuwen, J., Wiedermann, Jiří. Understanding Computation. A General Theory of Computational Processes. Utrecht: Utrecht University, 2019. 100 p. Technical Report, UU-CS-2019-012. ISSN 0924-3275.
Available: http://www.cs.uu.nl/research/techreps/repo/CS-2019/2019-012.pdf

[9] van Leeuwen, J., Wiedermann, Jiří. Turing Machines with One-sided Advice and Acceptance of the co-RE Languages. Fundamenta Informaticae. 2017, 153(4), 347-366. ISSN 0169-2968. Available: doi: 10.3233/FI-2017-1544

[10] van Leeuwen, J., Wiedermann, Jiří. Separating the Classes of Recursively Enumerable Languages Based on Machine Size. International Journal of Foundations of Computer Science. 2015, 26(6), 677-695. ISSN 0129-0541.
Available: doi: 10.1142/S0129054115500380

Focus 3. Evolutionary optimization strategies for automatic machine learning

Automatic machine learning (AutoML) procedures have recently acquired increasing attention. Our team has a long history of tackling AutoML problems with evolutionary optimization algorithms or related computational approaches. Particularly, the following results have been achieved in the evaluated period 2015-2019.

For global optimization tasks, the Covariance Matrix Adaptation Evolutionary Strategy (CMA-ES) represents a successful evolutionary continuous black-box optimizer. We developed a surrogate modeling technique for CMA-ES [11], based on a doubly trained surrogate regression model. In an extensive comparison, it achieved for some frequently occurring conditions a better performance than the so-far most successful surrogate models based on polynomial regression and support vector regression. Moreover, the algorithm won the 1. prize in the category Expensive black-box optimization of the optimization challenge accompanying the GECCO conference.

The evolution strategy of deep neural networks is a step towards an automatic architecture design. The automatic methods for architecture selection are still missing and neural architecture search (NAS) is a hot topic in neural network community. An algorithm for an optimization of a network architecture based on evolution strategies has been proposed [12]. The algorithm is inspired and designed directly for the Keras library which is one of the most common implementation tools for deep neural networks.

The unified framework for automatic learning of kernel networks with one hidden layer has been introduced [13]. The novel part includes the learning procedure that consists of meta-parameter tuning wrapping the standard parameter optimization part. Meta-parameter optimization improves the model performance significantly.

Another result concerns representation of machine learning workflow as direct acyclic graphs that can be efficiently generated and examined by strongly typed genetic programming approach. In the paper [14] we present a novel algorithm for workflow optimization which is implemented using the scikit-learn python ML library and tested on several problems from OpenML benchmark.

[11] Bajer, L., Pitra, Z., Repický, J., Holeňa, Martin. Gaussian Process Surrogate Models for the CMA Evolution Strategy. Evolutionary Computation. 2019, 27(4), 665-697. ISSN 1063-6560. Available: doi: 10.1162/evco_a_00244

[12] Vidnerová, Petra, Neruda, Roman. Evolution Strategies for Deep Neural Network Models Design. In: Hlaváčová, J., ed. Proceedings ITAT 2017: Information Technologies - Applications and Theory. Aachen & Charleston: Technical University & CreateSpace Independent Publishing Platform, 2017, 159-166. CEUR Workshop Proceedings, V-1885. ISBN 978-1974274741. ISSN 1613-0073.
Available: http://ceur-ws.org/Vol-1885/159.pdf

[13] Vidnerová, Petra, Neruda, Roman. Kernel Function Tuning for Single-Layer Neural Networks. International Journal of Machine Learning and Computing. 2018, 8(4), 354-360. ISSN 2010-3700. Available:
http://www.ijmlc.org/index.php?m=content&c=index&a=show&catid=79&id=831

[14] Křen, T., Pilát, M., Neruda, Roman. Automatic Creation of Machine Learning Workflows with Strongly Typed Genetic Programming. International Journal on Artificial Intelligence Tools. 2017, 26(5), 1760020. ISSN 0218-2130.
Available: doi: 10.1142/S021821301760020X

Focus 4. Statistical properties of methods and algorithms for data analysis and modeling

Statistical properties of machine learning methods and algorithms represent another focus of research of the team, including topics at the boundary between computer science and mathematical statistics.
Particularly, the research has been carried out in the following areas:

- Theoretical foundations of fuzzy sets [15, 16],
- Methods for multivariate (possibly high-dimensional) data analysis, including dimensionality reduction [17, 18, 19],
- Signal analysis based on nonparametric methodology [20],
- Hypothesis testing [21, 22].

The new class of radial fuzzy systems has been introduced [15]. The systems in this class use radial functions to implement membership functions of fuzzy sets and exhibit a shape preservation property in antecedents of their rules. The property is called the radial property. It enables the radial fuzzy systems to have their computational model mathematically tractable under both conjunctive and implicative representations of their rule bases. The coherence of radial implicative fuzzy systems was discussed and a sufficient condition on parameters of the system to exhibit coherent computation has been stated.

Two novel fuzzy measures and a modification of an existing fuzzy measure which are interaction-sensitive were presented in [16]: if classifier aggregation based on a fuzzy integral is used, they model not only the confidences of classifiers, but also their mutual similarities, contrary to the usually used Sugeno λ-measure that cannot model interactions between individual classifiers. The properties of the measures are first studied theoretically, and then the performance of the proposed measures is compared to the Sugeno λ-measure.

Fitting copulas to multidimensional data is an increasingly important method for analyzing dependencies, and the proposed quantifiers of observational calculus assess the results of estimating the structure of joint distributions of continuous variables by means of hierarchical Archimedean copulas. To this end, the existing theory of hierarchical Archimedean copulas has been slightly extended: It has been proven [17] that sufficient conditions for the function defining a hierarchical Archimedean copula to be indeed a copula, which have so far been rigorously established only for the special case of fully nested Archimedean copulas, hold in general. These conditions allow one to define three new generalized quantifiers, which are then thoroughly validated on four benchmark data sets and one data set from a real-world application.

A new supervised dimensionality reduction method is proposed [18], which is robust to the presence of outlying measurements in the data. It is suitable for high-dimensional data observed in two different groups (e.g. gene expressions over patients and over control persons). The method is based on a robust correlation coefficient, for which we prove the robustness and derive asymptotic properties, and a newly proposed regularized coefficient of multiple correlation. Numerical experiments reveal the new method to yield results, which are comparable with results of available (non-robust) methods over non-contaminated biomedical high-dimensional data. If the data are contaminated by severe noise or outliers, the new robust method yields results, which are superior to a variety of available methods.

Searching for an appropriate data representation in a space of fewer dimensions is the big challenge in large data set analysis. One of the most efficient methods to solve this task is factor analysis, which in the binary case is an NP-hard problem. Comparison of seven methods for Boolean factor analysis (BFA) in solving the so-called bars problem (BP) by means of information gain obtained in solving BP of different levels of complexity has allowed to reveal strengths and weaknesses of these methods [19]. It has been shown that the Likelihood maximization Attractor Neural Network with Increasing Activity (LANNIA), proposed by the team, is the most efficient BFA method in solving BP in most cases. Efficacy of the LANNIA method was shown as well, when applied to the real data from the Kyoto Encyclopedia of Genes and Genomes database, which contains full genome sequencing for 1368 organisms, and to text data set R52 (from Reuters 21578) typically used for categorizing labels.

It is known that particle filtering extends the capabilities of the Kalmán filter to the case when the distributions of the signal and observations are not Gaussian. The particle filter generates samples from the filtering distribution which are not i. i. d. (independent identically distributed) due to the resampling step in the particle filter's algorithm. In contrast, the standard kernel density estimation methodology is based on the i. i. d. assumption of available data. Hence the direct use of these methods in the particle filter may lead to inappropriate results. It was shown [20] that in spite of this obstacle the kernel estimates still converge to the true filtering densities for the class of Sobolev densities. Moreover, it presents the upper and lower bounds on the convergence rates in dependence on the order of the density and kernel used in the estimates. As these rates asymptotically coincide the kernel density estimates are optimal in the context of particle filtering.

According to the Neyman-Pearson fundamental lemma, the likelihood ratio test represents the most powerful statistical hypothesis test with a given level. However, it is vulnerable to the presence of measurement errors in the data. To overcome this vulnerability, we derive the likelihood ratio test also in the model with measurement errors to be the most powerful test, but with a modified significance level. As the main result, the minimax test reliable under measurement errors has been derived. The optimality of the test has been derived in the framework of robust optimization by means of an information-theoretical approach [21]. A numerical experiment reveals the novel test to be reliable under measurement errors in the data.

The statistical hypothesis tests based on so-called sequential ranks have been investigated [22]. Particularly, it is devoted to several different nonparametric tests in situations, when the data are observed in a sequential design, one measurement at a time. As the main results, several theorems about the optimality of tests based on sequential ranks are proven, which allow one to derive the locally most powerful sequential rank tests. Thus, some existing tests based on sequential ranks are proven to be optimal, while the new results bring the very first theoretical arguments in their favor, and also some new tests are proposed. Particular examples covered by the new theory include the task of comparing two random samples by a sequential version of the Wilcoxon test. A sequential testing procedure based on the investigated test statistics is also proposed.

[15] Coufal, David. Radial Fuzzy Systems. Fuzzy Sets and Systems. 2017, 319(15 July), 1-27. ISSN 0165-0114. Available: doi: 10.1016/j.fss.2016.06.012

[16] Štefka, D., Holeňa, Martin. Dynamic Classifier Aggregation using Interaction-Sensitive Fuzzy Measures. Fuzzy Sets and Systems. 2015, 270(1 July), 25-52. ISSN 0165-0114. Available: doi: 10.1016/j.fss.2014.09.005

[17] Holeňa, Martin, Bajer, L., Ščavnický, M. Using Copulas in Data Mining Based on the Observational Calculus. IEEE Transactions on Knowledge and Data Engineering. 2015, 27(10), 2851-2864. ISSN 1041-4347. Available: doi: 10.1109/TKDE.2015.2426705

[18] Kalina, Jan, Schlenker, Anna. A Robust Supervised Variable Selection for Noisy High-Dimensional Data. BioMed Research International. 2015, 320385. ISSN 2314-6133. Available: doi: 10.1155/2015/320385

[19] Frolov, A., Húsek, Dušan, Polyakov, P.Y. Comparison of Seven Methods for Boolean Factor Analysis and Their Evaluation by Information Gain. IEEE Transactions on Neural Networks and Learning Systems. 2016, 27(3), 538-550. ISSN 2162-237X. Available: doi: 10.1109/TNNLS.2015.2412686

[20] Coufal, David. Convergence rates of kernel density estimates in particle filtering. Statistics & Probability Letters. 2019, 153(October), 164-170. ISSN 0167-7152. Available: doi: 10.1016/j.spl.2019.06.013

[21] Broniatowski, M., Jurečková, J., Kalina, Jan. Likelihood ratio testing under measurement errors. Entropy. 2018, 20(12), 966. ISSN 1099-4300. Available: doi: 10.3390/e20120966

[22] Kalina, Jan. On Locally Most Powerful Sequential Rank Tests. Sequential Analysis. 2017, 36(1), 111-125. ISSN 0747-4946. Available: doi: 10.1080/07474946.2016.1275501

Focus 5. Applications of machine learning methods

Members of the team have rich experience with applications of machine learning methods to various practical tasks. In the evaluated period 2015-2019, they contributed to applications in various separation (classification) or prediction tasks overviewed in this section. Particular application results were achieved (often in cooperation with experts from the corresponding disciplines) in

- Nuclear physics [23, 24],
- Engineering and robotics [25, 26],
- Molecular genetics [27],
- Economics [28].

In cooperation with FERMILAB, USA, machine learning methods were used to evaluate D0 and NOvA experiment data, especially as supporting methods in analysis of some parameters of neutrino oscillation within the NOvA experiment cooperation group [23, 24].

In the field of material engineering of semiconductor chips, especially in directional solidification of silicon, the solid-liquid interface shape plays a crucial role for the quality of crystals. The interface shape can be influenced by forced convection using travelling magnetic fields. Up to now, there has been no general and explicit methodology to identify the relation and the optimum combination of magnetic and growth parameters e.g., frequency, phase shift, current magnitude and interface deflection in a buoyancy regime. 2D CFD modeling was used [25] to generate data for the design and training of artificial neural networks and for Gaussian process modeling. The aim was to quickly assess the complex nonlinear dependencies among the parameters and to optimize them for the interface flattening. The first encouraging results are presented and the pros and cons of artificial neural networks and Gaussian process modeling discussed.

Control of multi-body systems in both robots and humans may face the problem of destabilizing dynamic coupling effects arising between linked body segments. The state-of-the-art solutions in robotics are full state feedback controllers. For human hip-ankle coordination, a more parsimonious and theoretically stable alternative to the robotics solution, in terms of the Eigenmovement (EM) control, has been suggested [26]. EM are kinematic synergies designed to describe the system with several degrees of freedom, and its control, with a set of independent, and hence coupling-free, scalar equations. Tests conducted in the "posture control laboratory" show that the EM controller provides stable humanoid robot control with proactive ("voluntary") movements and reactive balancing of positions during tilting and translating support surfaces. Although a preliminary robot-human comparison reveals similarities and differences, authors conclude that the EM concept is a valid candidate for HR control.

The causes of errors that degrade standard tools for analyzing image information obtained through microchips in molecular genetics have been studied [27]. To eliminate these errors a robust procedure able to detect a greater percentage of outlying values in the data versus other commonly used methodologies was designed. This resulted in better subsequent classification of tasks analyzing specific cardiovascular genetic studies.

Tourist service infrastructure across 141 countries has been modeled [28] as a response of 12 characteristics of the Travel and Tourism Competitiveness Index (TTCI), reported by the World Economic Forum. In the connection with this topic we discuss various approaches to regression modelling for heteroscedastic multicollinear data, with a focus on regularized regression quantiles. As the main result, it was shown that regression quantiles and their lasso estimates turn out to be more suitable than more traditional econometric tools.

[23] Acero, M. A., Adamson, P., Aliaga, L., Filip, P., Hakl, František, Lokajíček, M., Zálešák, J. Observation of seasonal variation of atmospheric multiple-muon events in the NOvA Near Detector. Physical Review D. 2019, 99(12), 122004. ISSN 2470-0010. Available: doi: 10.1103/PhysRevD.99.122004

[24] Acero, M. A., Adamson, P., Aliaga, L., Filip, P., Hakl, František, Lokajíček, M., Zálešák, J. First measurement of neutrino oscillation parameters using neutrinos and antineutrinos by NOvA. Physical Review Letters. 2019, 123(15), 151803. ISSN 0031-9007. Available: doi: 10.1103/PhysRevLett.123.151803

[25] Dropka, N., Holeňa, Martin. Optimization of Magnetically Driven Directional Solidification of Silicon Using Artificial Neural Networks and Gaussian Process Models. Journal of Crystal Growth. 2017, 471(1 August), 53-61. ISSN 0022-0248. Available: doi: 10.1016/j.jcrysgro.2017.05.007

[26] Alexandrov, A.V., Lippi, V., Mergner, T., Frolov, A., Hettich, G., Húsek, Dušan. Human-Inspired Eigenmovement Concept Provides Coupling-Free Sensorimotor Control in Humanoid Robot. In: Sartori, M., Valero-Cuevas, F. J., Schouten, A. C., Tresch, M., Nakamura, Y., Sreenivasa, M., eds. Neuromechanics and Control of Physical Behavior: from Experimental and Computational Formulations to Bio-inspired Technologies. Lausanne: Frontiers, 2019, 208-220. Frontiers Research Topics. ISBN 978-2-88945-946-9. Available: doi: 10.3389/fnbot.2017.00022

[27] Kalina, Jan. A Robust Pre-processing of BeadChip Microarray Images. Biocybernetics and Biomedical Engineering. 2018, 38(3), 556-563. ISSN 0208-5216. Available: doi: 10.1016/j.bbe.2018.04.005

[28] Kalina, Jan, Vašaničová, P., Litavcová, E. Regression Quantiles under Heteroscedasticity and Multicollinearity: Analysis of Travel and Tourism Competitiveness. Ekonomický časopis. 2019, 67(1), 69-85. ISSN 0013-3035. Available: https://www.sav.sk/journals/uploads/0125113201%2019%20Kalina%20+%20RS.pdf

Research activity and characterisation of the main scientific results

I. Main Scientific Results

In this section we go over the main scientific results achieved by the team during the evaluated period. All the results discussed here were published in impacted scientific journals during the evaluated period.

a) Basic research in mathematical statistics

One area of the team's activity involves basic research in mathematical statistics, which serves as a foundation for the research methodology in statistical data science and modelling. One important research direction is exemplified by studying heavy tailed distributions, such as e.g. Cauchy distribution, which may not be characterised as usual in terms of the location and scale parameter [1]. In this case the centrality parameter, the mean, may not exist which may cause a problem in providing statistical inference based on such probability distributions. Apart from the mean, mode, and median, we obtained additional new characteristic of the central tendency of continuous distributions. The mean of heavy-tailed distributions as well as the mode of distributions with partial support may not exist, and the median may not be an eligible description. The proposed score mean exists for any transformed distribution with unimodal prototype and we suggested it as the best characteristic of central tendency of all continuous distributions. The main result is that sums of score random variables obey the central limit theorem and can be used for directly estimating the score mean without the use of any current estimator based on the moment or maximum likelihood method. We also provided new estimators of the extreme value index for heavy tailed distributions based on transformed score (t-score). The new t-Hill estimator [2] estimates the extreme value index of a distribution function with a regularly varying tail. In cases of contamination with heavier tails than those of the original sample, we have shown that t-Hill outperforms several robust tail estimators, especially in small samples. A simulation study emphasised the fact that the level of contamination is playing a crucial role. The larger the contamination, the better are the t-score moment estimates.

We studied processes for estimating multilevel maximum likelihood with application to covariance matrices in the context of data assimilation in meteorology and oceanography, where the dimension of state vector describing the atmosphere or ocean is in the order of millions or larger [3]. Advantages of using low-parametric models for estimating covariance matrix using a small sample were pointed out and a large class of estimating methods for high-dimensional covariance matrices was considered, namely the shrinkage estimators. The latter class of estimators was also researched in analysing sparse data from genome-wide association studies [4] and found very beneficial in improving the classification properties of linear and quadratic discriminant analysis in the context of sparse data, i.e. comparatively low number of records in comparison with very large number of covariates (e.g. genes).

New change point methods were proposed for weakly correlated observations in time series which rely on ratio type test statistics based on maxima of the cumulative sums [5]. These detection procedures for the abrupt change in mean were also robustified by considering a general score function. The main advantage of the proposed approach is that the variance of the observations neither has to be known nor estimated. The block bootstrap method was developed in order to obtain better approximations for the test's critical values. We further proposed new statistical procedures for a change in the means problem within a very general panel data structure. It allowed for mutually dependent panels, unequal variances across the panels, and, possibly, a very short follow up period. Two competitive ratio type test statistics were introduced, and their asymptotic properties were derived for a large number of available panels. The proposed tests proved to be consistent and their empirical properties were investigated in an extensive simulation study.

[1] Fabián, Zdeněk. Mean, mode or median? The score mean. Communications in Statistics - Theory and Methods. 2019, Online: 12 December 2019. ISSN 0361-0926. Available: doi: 10.1080/03610926.2019.1666142

[2] Jordanova, P., Fabián, Zdeněk, Hermann, P., Střelec, L., Rivera, A., Girard, S., Torres, S., Stehlík, M. Weak Properties and Robustness of t-Hill Estimators. Extremes. 2016, 19(4), 591-626. ISSN 1386-1999. doi: 10.1007/s10687-016-0256-2

[3] Turčičová, Marie, Mandel, Jan, Eben, Kryštof. Multilevel maximum likelihood estimation with application to covariance matrices. Communications in Statistics - Theory and Methods. 2019, 48(4), 909-925. ISSN 0361-0926. Available: doi: 10.1080/03610926.2017.1422755

[4] Valenta, Zdeněk, Kalina, Jan. Exploiting Stein's Paradox in Analysing Sparse Data from Genome-Wide Association Studies. Biocybernetics and Biomedical Engineering. 2015, 35(1), 64-67. ISSN 0208-5216. Available: doi: 10.1016/j.bbe.2014.10.004

[5] Peštová, Barbora, Pešta, M. Abrupt change in mean using block bootstrap and avoiding variance estimation. Computational Statistics. 2018, 33(1), 413-441. ISSN 0943-4062. Available: doi: 10.1007/s00180-017-0785-4

b) Interdisciplinary research

The team's involvement in interdisciplinary research contributes to making scientific progress in several scientific disciplines. Here the focus is on proposing statistical designs appropriate for making intended inferences, data protocols specification and adaptation of suitable statistical methodology. The team's contribution may be seen in bringing advanced know-how and adaptive solutions to other scientific disciplines without which drawing proper inferences and knowledge would not be possible. That, of course, requires a deep comprehension and grasp of the scientific topic at hand and its translation into the statistical realm.

1. Neurosciences

Interdisciplinary research in the field of neuroscience represents an important part of the scientific collaboration conducted by the team, which allows for bringing state-of-the-art statistical methodology to neuroscience and helping in identifying important determinants of non-communicable civilisation diseases affecting the normal functioning of the brain.

Repeated-measures ANOVA with rarely used features helped us to uncover strong effects of prenatal exposure to methamphetamine on the system [6] even when the exposure was limited to prenatal phase. Toxicological effects have longer lasting effects as currently considered and seem to affect the excitatory-inhibitory balance in the brain having strong implications for cognitive and behavioural functioning.

The team members studied aging and chronic sleep deprivation as well-recognised risk factors for Alzheimer's disease (AD) [7]. They helped to determine aging and sleep deprivation as potential factors responsible for cognitive decline in the elderly, affecting the pathobiology of AD and the neurodegenerative process. The team members helped in identifying a combination of prenatal and postnatal methamphetamine exposure as factors increasing the risk of dopaminergic deficits via alterations in the activity of surface-expressed dopamine transporters, especially in preadolescent females. They have shown a high sensitivity and specificity of a newly proposed biomarker, intracellular amyloid-beta peptide-binding protein, to Alzheimer's disease.

In another study [8], the team members helped to determine that surface tension values are not a suitable biomarker for AD, but the ratio of ThioflavinT-based fluorescence to intrinsic amyloid fluorescence in cerebrospinal fluid does appear to be an acceptable supportive diagnostic biomarker for Alzheimer´s Disease.

The study of children with specific language impairments [9] confirmed previous results indicating deficient connectivity of the arcuate fascicle. An abnormal development of a ventral language stream in children with specific language impairments was newly indicated, which may have clinical implications for paediatric patients undergoing brain surgery.

The team members participated in the design and led the analysis of the long-term project Multicentre European Multiple Sclerosis Survey [10], which was a collaboration of researchers from more than 15 European countries. They studied the effect of physiotherapeutic services on the MS from the organisational point of view. Study [11], led by one team member, employed a unique cluster analysis to better explain regional differences in the use of physical therapy interventions, helping to understand the reasons why certain interventions are or are not applied.

In a study of dopamine agonist-resistant prolactinomas [12], a team member helped in confirming the effectivity of Gamma knife radiosurgery. Normo-prolactinaemia was achieved in the majority of patients, either after discontinuation of, or while continuing to take, dopamine agonists. Tumour growth was arrested in all cases. The risk of developing hypopituitarism may be limited by regularly administrating safe doses to the pituitary and infundibulum.

[6] Fujáková-Lipski, M., Kaping, D., Šírová, J., Horáček, J., Páleníček, T., Zach, P., Klaschka, Jan, Kačer, P., Syslová, K., Vrajová, M., Bubeníková-Valešová, V., Beste, D., Šlamberová, R. Trans-generational Neurochemical Modulation of Methamphetamine in the Adult Brain of the Wistar Rat. Archives of Toxicology. 2017, 91(10), 3373-3384. ISSN 0340-5761. doi: 10.1007/s00204-017-1969-y

[7] Krištofiková, Z., Šírová, J., Klaschka, Jan, Ovsepian, S. V. Acute and Chronic Sleep Deprivation-Related Changes in N-methyl-D-aspartate Receptor—Nitric Oxide Signalling in the Rat Cerebral Cortex with Reference to Aging and Brain Lateralization. International Journal of Molecular Sciences. 2019, 20(13), 3273. E-ISSN 1422-0067. Available: doi: 10.3390/ijms20133273

[8] Gažová, Z., Bednáriková, Z., Bartoš, A., Klaschka, Jan, Krištofíková, Z. Surface Tension and Intrinsic Amyloid Fluorescence of Serum and Cerebrospinal Fluid Samples in Alzheimer´s Disease. Biomarkers in Medicine. 2019, 13(4), 267-277. ISSN 1752-0363. Available: doi: 10.2217/bmm-2018-0319

[9] Vydrová, R., Komárek, V., Šanda, J., Štěrbová, K., Jahodová, A., Maulisová, A., Žáčková, J., Reissigová, Jindra, Kršek, P., Kyncl, M. Structural Alterations of the Language Connectome in Children with Specific Language Impairment.Brain and Language. 2015, 151(December), 35-41. ISSN 0093-934X.
Available: doi: 10.1016/j.bandl.2015.10.003

[10] Řasová, K., Freeman, J., Martinková, Patrícia, Pavlíková, M., Cattaneo, D., Jonsdottir, J., Henze, T., Baert, I., Van Asch, P., Santoyo, C., Smedal, T., Beiske, A.G., Stachowiak, M., Kovalewski, M., Nedeljkovic, U., Bakalidou, D., Guerreiro, J. M. A., Nilsagard, Y., Dimitrova, E. N., Habek, M., Armutlu, K., Donzé, C., Ross, E., Ilie, A. M., Martic, A., Romberg, A., Feys, P. The Organisation of Physiotherapy for People with Multiple Sclerosis Across Europe: A Multicentre Questionnaire Survey. Bmc Health Services Research. 2016, 16(1), 552. ISSN 1472-6963.
Available: doi: 10.1186/s12913-016-1750-6

[11] Martinková, Patrícia, Freeman, J., Drabinová, Adéla, Erosheva, E., Cattaneo, D., Jonsdottir, J., Baert, I., Smedal, T., Romberg, A., Feys, P., Alves-Guerreiro, J., Habek, M., Henze, T., Santoyo Medina, C., Beiske, A., Van Asch, P., Bakalidou, D., Salci, Y., Dimitrova, E.N., Pavlíková, M., Řasová, K. Physiotherapeutic interventions in multiple sclerosis across Europe: regions and other factors that matter. Multiple Sclerosis and Related Disorders. 2018, 22(May), 59-67. ISSN 2211-0348.
Available: doi: 10.1016/j.msard.2018.03.005

[12] Ježková, J., Hána, V., Kosák, M., Kršek, M., Liščák, R., Vymazal, J., Pecen, Ladislav, Marek, J. Role of gamma knife radiosurgery in the treatment of prolactinomas. Pituitary. 2019, 22(4), 411-421. ISSN 1386-341X.
Available: doi: 10.1007/s11102-019-00971-x

2. Cardiovascular research

In a study of aspirin use in primary prevention of cardiovascular disease [13] the team contributed to finding that the drug reduced relative risk of myocardial infarction and ischaemic stroke. On the other hand, the modelling revealed an association with an increased rate of major bleeding events. The team members helped in demonstrating that aspirin use did not translate into a net clinical benefit, when adjusted for event-associated mortality risk, and did not reduce all-cause or cardiovascular mortality.

Further research concerned thromboembolic and bleeding risks in patients with heart failure (HF) and atrial fibrillation (AF) [14]. The team members contributed to finding that the HF subtype predicts thromboembolic risks in mostly anticoagulated patients with AF. Age, HF type, and NYHA class were established as independent predictors of thromboembolic events. Major bleeding did not differ significantly between the groups.

Another study considered optical coherence tomography (OCT) guidance during stent implantation in primary percutaneous coronary intervention [15]. The team members contributed to demonstrating the safety and possible merit of OCT-guided second-generation DES stent deployment during primary percutaneous coronary intervention (pPCI) in STEMI patients.

Using the data patterns of coronary heart disease (CHD) mortality rates in the Czech Republic over the past 50 years we attempted to derive short-term predictions of CHD mortality [16]. Age-standardised mortality rates were predicted to decrease in 2015–2019, crude mortality rates were expected to increase due to an increase in average life expectancy. The burden of deaths was expected to move to the age group of 85 years and older, primarily in female subjects.

[13] Gelbenegger, G., Postula, M., Pecen, Ladislav, Halvorsen, S., Lesiak, M., Schoergenhofer, C., Jilma, B., Hengstenberg, C., Siller-Matula, J. M. Aspirin for primary prevention of cardiovascular disease: A meta-analysis with a particular focus on subgroups. BMC Medicine. 2019, 17(1), 198. ISSN 1741-7015.

Available: doi: 10.1186/s12916-019-1428-0

[14] Siller-Matula, J. M., Pecen, Ladislav, Patti, G., Lucerna, M., Kirchhof, P., Lesiak, M., Huber, K., Verheugt, F. W. A., Lang, I., Renda, G., Schnabel, R. B., Wachter, R., Kotecha, D., Sellal, J. - M., Rohla, M., Ricci, F., De Caterina, R. Heart Failure Subtypes and Thromboembolic Risk in Patients with Atrial Fibrillation: The PREFER in AF - HF substudy. International Journal of Cardiology. 141-147. ISSN 0167-5273. Available: doi: 10.1016/j.ijcard.2018.04.093

[15] Kala, P., Červinka, P., Jakl, M., Kanovský, J., Kupec, A., Špaček, R., Kvasnak, M., Poloczek, M., Červinková, M., Bezerra, H., Valenta, Zdeněk, Attizzani, G., Schnell, A., Hong, L., Costa, M. OCT guidance during stent implantation in primary PCI: A randomized multicenter study with nine months of optical coherence tomography follow-up. International Journal of Cardiology. 2018, 250(1 January), 98-103. ISSN 0167-5273. Available: doi: 10.1016/j.ijcard.2017.10.059

[16] Reissigová, Jindra, Zvolský, M. Short-term Prediction of Coronary Heart Disease Mortality in the Czech Republic Based on Data from 1968-2014. Central European Journal of Public Health. 2018, 26(1), 10-15. ISSN 1210-7778. Available: doi: 10.21101/cejph.a4936

3. Physiology

The team members contributed to describing the development and validation of homeostasis concept inventory [17] and introduced educational tests-based procedure for understanding the concept of homeostasis, detailed psychometric analysis for checking test validity, reliability and functioning of individual test items which brings the process of test validation to a higher level.

In another effort, the team members helped in the development and validation of a conceptual framework for the core concept of cell-cell communication [18]. The concept extends to every function of the nervous and endocrine system in regulation of different body systems, includes synaptic, paracrine, and endocrine mechanisms of communication, and integrates molecular, cellular, tissue, organ, and organismal aspects of communication and regulation across a range of levels of biological organisation.

[17] McFarland, J. L., Price, R. M., Wenderoth, M. P., Martinková, Patrícia, Cliff, W., Michael, J., Modell, H., Wright, A. Development and Validation of the Homeostasis Concept Inventory. CBE-Life Sciences Education. 2017, 16(2), ar35. ISSN 1931-7913. Available: doi: 10.1187/cbe.16-10-0305

[18] Michael, J., Martinková, Patrícia, McFarland, J. L., Wright, A., Cliff, W., Modell, H., Wenderoth, M. P. Validating a Conceptual Framework for the Core Concept of "Cell-cell Communication". Advances in Physiology Education. 2017, 41(2), 260-265. ISSN 1043-4046. Available: doi: 10.1152/advan.00100.2016

4. Education research

The psychometrics group established the importance and superiority of differential item functioning (DIF) analysis over the traditional methods for the detection of between-group differences in the development and validation of educational tests [19] and introduced new methods for the detection of DIF in presence of guessing [20] and in longitudinal designs [21].

The group also contributed to analysing disparities in ratings of internal and external applicants to teaching positions using applicant data from Spokane Public Schools [22]. We tested for biases in rating while accounting for measures of teacher applicant qualifications and quality and proposed model-based inter-rater reliability (IRR) estimates that allows for accounting of various sources of measurement error, the hierarchical structure of the data, and to test whether covariates, such as applicant status, moderate IRR. Applicants external to the district typically obtained lower ratings for job applications, as compared with internal applicants. This gap in ratings remained significant even after including measures of qualifications and qualities such as experience, state licensure scores, or estimated teacher added value.

A wide range of psychometric methods, including the methods newly proposed by the group, are available in the R package and interactive application ‘ShinyItemAnalysis’. In the companion paper describing the software [23], the group argued that complex psychometric analysis should become a routine part of test development in order to gather proofs of reliability and measurement validity.

[19] Martinková, Patrícia, Drabinová, Adéla, Liaw, Y. L., Sanders, E. A., McFarland, J. L., Price, R. M. Checking Equity: Why Differential Item Functioning Analysis Should Be a Routine Part of Developing Conceptual Assessments. CBE-Life Sciences Education. 2017, 16(2), rm2. ISSN 1931-7913.
Available: doi: 10.1187/cbe.16-10-0307

[20] Martinková, Patrícia, Goldhaber, D., Erosheva, E. Disparities in Ratings of Internal and External Applicants: A Case for Model-based Inter-rater Reliability. PLoS ONE. 2018, 13(10), e0203002. E-ISSN 1932-6203.
Available: doi: 10.1371/journal.pone.0203002

[21] Drabinová, Adéla, Martinková, Patrícia. Detection of Differential Item Functioning with Nonlinear Regression: A Non-IRT Approach Accounting for Guessing. Journal of Educational Measurement. 2017, 54(4), 498-517. ISSN 0022-0655.
Available: doi: 10.1111/jedm.12158

[22] Martinková, P., Hladká, Adéla, Potužníková, E. Is academic tracking related to gains in learning competence? Using propensity score matching and differential item change functioning analysis for better understanding of tracking implications. Learning and Instruction. 2019, Online 3 December 2019, 101286. ISSN 0959-4752.
Available: doi: 10.1016/j.learninstruc.2019.101286

[23] Martinková, Patrícia, Drabinová, Adéla. ShinyItemAnalysis for Teaching Psychometrics and to Enforce Routine Analysis of Educational Tests. R Journal. 2018, 10(2), 503-515. ISSN 2073-4859. doi: 10.32614/RJ-2018-074

5. Ecology, Climatology and Environmental Research

The team members contributed to studying spatial congruence between the richness of ancient forest species and other forest-dwelling groups and provided evidence for a positive relationship between the two entities in temperate European forests [24]. They helped in concluding that retention of small mutually isolated refuges of endangered species interspersed within large areas of managed forests may be crucial for the mitigation of biodiversity decline.

They further contributed to finding that a thermal gradient from the forest edge to the forest interior appears to be a significant modulator of climate change which affects substantial areas of temperate forest remnants dispersed throughout agricultural landscapes in Central Europe [25].

In another study [26], using data from the Carpathian Mountains the team members contributed to finding that basal area of late-successional Picea abies forests varied across a range of spatial scales, with climatic drivers being most important at coarse scales and natural disturbances acting as the primary driver of forest heterogeneity at fine scales. The findings suggest that warming could increase the basal area of northern sites, but potential increasing disturbances could disrupt these environmental responses.

Statistical modelling helped in identifying the extent of variations in the severity of disturbances, within which Picea abies forests are able to regenerate and retain their monospecific character [27]. That is increasingly relevant as disturbance regimes continue to shift under global climate change. It was found that temperate spruce forests persist in having a monospecific character, despite variations in the severity of disturbances, due to strategic advantages they enjoy in light of advanced regeneration.

In a study of fog occurrence [28], the group members inquired about possible drivers of changes in the probability with which systematic fogs occur. The approach based on a flexible semiparametric model, acknowledging major structural types of fog presence and possibly complex covariate effects, identified relative humidity and air pollutants as the most important factors explaining the probability of fog occurrence.

Another study [29] explored the feasibility of retrieving effective cloud field parameters from systematic radiometric high frequency measurements. Having proposed two novel statistical procedures we concluded that the relationship between cloud shade and point cloudiness computed using estimated cloud field parameters recovers the typical relationship derived from the measurements.

As a novelty in sunshine duration analysis, the team members studied the duration using a time-to-event approach [30]. They discovered that the increase of extra-terrestrial solar radiation tends to elevate the risk of stopping any clear or dark interval. Stochastically shorter intervals were found to be associated with increasing sun altitude.

The team members further studied the association between the prevalence of childhood respiratory diseases and long-term exposure to air pollution in the burdened Moravian–Silesian region of the Czech Republic [31]. The association between traffic indicators and respiratory health was not established. Long-term exposure to air pollution was found to be associated with childhood allergic rhinitis and the deterioration of respiratory symptoms.

The team members also contributed to introducing a new data analysis tool, the linear score regression. This methodology was applied on ecologically and societally important small samples from administrative water quality data in Arica and Parinacota in Chile [32]. The Pareto distribution of the right tail of residuals was justified and a dynamic system was developed explaining why generic tests for zero slope regression do not reject the null hypothesis.

[24] Hofmeister, J., Hošek, J., Brabec, Marek, Hermy, M., Dvořák, D., Fellner, R., Malíček, J., Palice, Z., Tenčík, A., Holá, E., Novozámská, E., Kuras, T., Trnka, F., Zedek, M., Kašák, J., Gabris, R., Sedláček, O., Tajovský, K., Kadlec, T. Shared affinity of various forest-dwelling taxa point to the continuity of temperate forests. Ecological Indicators. 2019, 101(Jun 2019), 904-912. ISSN 1470-160X.
Available: doi: 10.1016/j.ecolind.2019.01.018

[25] Hofmeister, J., Hošek, J., Brabec, Marek, Střalková, R., Mýlová, P., Bouda, M., Pettit, J., Rydval, M., Svoboda, M. Microclimate Edge Effect in Small Fragments of Temperate Forests in the Context of Climate Change. Forest Ecology and Management. 2019, 448(15 September), 48-56. ISSN 0378-1127.
Available: doi: 10.1016/j.foreco.2019.05.069

[26] Janda, P., Tepley, A. J., Schurman, J. S., Brabec, Marek, Nagel, T.A., Bače, R., Begovič, K., Chaskovskyy, O., Čada, V., Dušátko, M., Frankovič, M., Kameniar, O., Kozák, D., Lábusová, J., Langbehn, T., Málek, J., Mikoláš, M., Nováková, M. H., Svobodová, K., Synek, M., Teodosiu, M., Trotsiuk, V., Svoboda, M. Drivers of basal area variation across primary late-successional Picea abies forests of the Carpathian Mountains. Forest Ecology and Management. 2019, 435(1 March), 196-204. ISSN 0378-1127. Available: doi: 10.1016/j.foreco.2018.12.045

[27] Bače, R., Schurman, J.S., Brabec, Marek, Čada, V., Deprés, T., Janda, P., Lábusová, J., Mikoláš, M., Morrissey, R. C., Mrhalová, H., Nagel, T.A., Nováková, M. H., Seedre, M., Synek, M., Trotsiuk, V., Svoboda, M. Long-term Responses of Canopy-understorey Interactions to Disturbance Severity in Primary Picea abies Forests. Journal of Vegetation Science. 2017, 28(6), 1128-1139.
ISSN 1100-9233. Available: doi: 10.1111/jvs.12581

[28] Hůnová, I., Brabec, Marek, Malý, Marek, Valeriánová, A. Revisiting Fog as an Important Constituent of the Atmosphere. Science of the Total Environment. 2018, 636(15 September), 1490-1499. ISSN 0048-9697.
Available: doi: 10.1016/j.scitotenv.2018.04.322

[29] Paulescu, M., Badescu, V., Brabec, Marek. Retrieval of effective cloud field parameters from radiometric data. Theoretical and Applied Climatology. 2018, 133(1-2), 437-446. ISSN 0177-798X. Available: doi: 10.1007/s00704-017-2192-8

[30] Brabec, Marek, Dumitrescu, A., Paulescu, M., Badescu, V. A new perspective on the sunshine duration variability. Theoretical and Applied Climatology. Published online: 28 November 2019. ISSN 0177-798X. doi: 10.1007/s00704-019-03046-w

[31] Puklová, V., Žejglicová, K., Kratěnová, J., Brabec, Marek, Malý, Marek. Childhood respiratory allergies and symptoms in highly polluted area of Central Europe. International Journal of Environmental Health Research. 2019, 29(1), 82-93.
ISSN 0960-3123. Available: doi: 10.1080/09603123.2018.1514458

[32] Stehlík, M., Núñez Soza, L., Fabián, Zdeněk, Jiřina, Marcel, Jordanova, P., Arancibia, S. C., Kiselák, J. On Ecological Aspects of Dynamics for Zero Slope Regression for Water Pollution in Chile. Stochastic Analysis and Applications. 2019, 37(4), 574-601. ISSN 0736-2994. Available: doi: 10.1080/07362994.2019.1592692

6. Ethology

In the context of animal behavioural sciences, the team members studied the potential for the use of marginal generalised linear modelling of correlated data, as an alternative to using generalized mixed-effect models [33]. Using four examples from behavioural research they demonstrated the use, advantages and limits of generalised estimating equations approach, as implemented in the ‘geepack’ package in R.

[33] Pekár, S., Brabec, Marek. Generalized estimating equations: A pragmatic and flexible approach to the marginal GLM modelling of correlated data in the behavioural sciences. Ethology. 2018, 124(2), 86-93. ISSN 0179-1613.
Available: doi: 10.1111/eth.12713

c) Monographs

The two monographs present classes of generalised linear and non-linear statistical models, and their deployment in the environment of the statistical analytic system R. They serve as a practical companion for data analysis using R and involve worked examples chosen to demonstrate a correct approach to model formulation and note frequent errors and problems occurring in data analysis. The texts encourage statistical thinking about problems without requiring a deep mathematical background.

[34] Pekár, S., Brabec, Marek. Modern Analysis of Biological Data. Generalized Linear Models in R. Brno: Masarykova univerzita, 2016. ISBN 978-80-210-8019-5. Available: http://aleph.nkp.cz/F/?func=direct&doc_number=002785092&local_base=NKC

[35] Brabec, Marek, Pekár, S. Moderní analýza biologických dat 3. Nelineární modely v prostředí R. Brno: Masarykova univerzita, 2019. ISBN 978-80-210-9277-8

II. Research for Practice – applied research

a) Statistical modelling for the Czech Ministry of Finance

During the evaluated period the team participated in several projects involving applied research for practice. The Czech Ministry of Finance involved our team members in a major project developing the models for monitoring financial flow in state-supervised institutions. We developed spatiotemporal statistical models for the structured analysis of financial indicators of various public institutions. Modelling involved a-priori knowledge solicited from financial and accounting specialists, proceeded in both static and dynamic variants, involving autoregressive analysis on different spatial scales of interest. Major results were published in

[36] Brabec, Marek, Malý, Marek. Časovo-prostorová statistická analýza finančních dat. Forum Statisticum Slovacum. 2019, 15(1), 1-9. ISSN 1336-7420
Available: http://www.ssds.sk/casopis/archiv/2019/fss0119.pdf#page=3

b) Statistical calibration of numerical weather forecast models

Modelling energy exchanges between buildings and their outdoor environment in Prague 6 provided important results directly useable in building design and urbanism. Details of building façade choices may influence the air flow and temperature in street canyons, related to the comfort and well-being of pedestrians. Team members performed calibrations for predicting solar irradiation in renewable energy management. Our work on flexible semiparametric calibration involved international cooperation with MeteoRomania which helped improving the Romanian ALARO weather prediction model.

c) Statistical methods for vibrodiagnostics

This interdisciplinary collaboration between two academic institutions (ICS CAS and the Institute of Thermomechanics CAS) focused on the improvement of blade tip timing of turbines at nuclear power plants. The team members formulated a generalised additive model for the time-varying spectral components obtained from large scale vibrodiagnostics data. Due to simultaneous deployment of several sensors, the Nyquist frequency was increased to study both asynchronous and synchronous frequencies at both constant and time-varying coefficient regimes. The fact that several frequencies could be estimated in one coherent statistical model is important for formalised statistical inference and the study of background spectral properties.

d) Modelling uncertainty in thermodynamic calculations, with applications in Energy

The team members attempted to synergise competences across the disciplines of mathematical statistics (ICS CAS) and thermophysical measurement theory (Institute of Thermomechanics CAS) to help develop a tool for the assessment of uncertainty of estimates of physical variables and constants derived from the data obtained in complex measurement campaigns. They experimented with both the systematic part of statistical model of various complexity (e.g., linear, nonlinear, nonparametric) and the detailed probabilistic properties of random error (e.g., distributional properties, random effects, partially multiplicative effects, etc.).
Working with complicated models, the team began with frequentist inference and then tested the potential for Bayesian inference. The progress was discussed at regular interdisciplinary workshops involving the ICS CAS and IT CAS.

e) Detection of defects in the gas pipeline operated by the NET4GAS company

In 2019, the task of using sample surveys and statistical quality control to detect faults in the NET4GAS gas pipeline based on current inspection diggings was explored. The occurrence of defects appeared to be non-homogeneous, and, due to incomplete data it was not possible to assess the probability of fault occurrence in statistical terms. The study continues with analysing supplemented and extended data files.

f) Statistical methods for psychometric analysis of behavioural measurements

The psychometrics group developed two software packages for analysing educational, psychological, health-related and other behavioural measurement data: ShinyItemAnalysis [22] and difNLR [20]. The latter was evaluated in a national evaluation of public research institutions and obtained the highest grade (world-leading), obtained by only 25 out of 6000 submitted non-bibliometric results.

Research activity and characterisation of the main scientific results

The search for order in chaos entails an effort to understand the complex behaviour of real-world phenomena, recorded in time-dependent experimental data and time series. In this endeavour the following research directions can be specified.

Nonlinear dynamics and causality

"More is different," this simple sentence of the Nobel laureate P. W. Anderson reflects the complex reality in which the behaviour of complex systems, consisting of many interacting elements, cannot be explained by a simple extrapolation of the laws describing the behaviour of a few elements. In order to understand complex systems, one needs to uncover interactions of their elements in the form of coupling, information transfer, or synchronization. This research direction has solid roots in previous results of the team: Milan Paluš and collaborators proposed the concept of synchronization as an adjustment of information rates and the information-theoretic approach to **detect causality** (in the Granger sense) from time series. The standard-setting papers in this area have gained international recognition and more than a thousand citations. In a recent work, the performance of different causality methods has been evaluated together with the possibility to infer information transfer delay (the time needed for transfer of information from the cause to the effect variable). A novel systematic generalization of standard Granger causality to nonlinear stochastic processes has been proposed. The behaviour of causality detection methods in the case of time reversal has been studied. Intensive research activity followed the original idea of cross-scale causality, including the development of methods and applications in climate science such as dynamics of temperature in European mid-latitudes and of El Niño.

Network inference and analysis

The dynamics of complex systems is frequently reflected in measurements of many variables or a few variables in many spatial locations. Therefore, multidimensional data reduction and clustering is also one of the research topics of this team. The work in this area included the development and application of methods for **studying causal networks** in large systems, the study of topological properties of correlation networks, development and application of methods to characterize succinctly the structure of interactions in climate as well as **cybersecurity applications**, methods for quantification of higher-order interactions going beyond standard network representation, dealing with methodological challenges related to data nonstationarity and network inference as well as a set of related problems in graph theory.

Neuroimaging data analysis and modelling

One of the key areas of application of complex systems research is the development and application of tailored methods for **analysis of brain activity**. The team has worked on novel methods to preprocess and analyse brain activity data, and contributed to a range of experimental studies. Apart from this data-driven approach, the team is involved in devising and analysing the behaviour of realistic **dynamical models of brain activity**. One of the most fascinating areas of study concerns the phenomena emergent from complex dynamics in neuroscience. The studied questions include: What is the mechanism of emergence of epileptic seizure from normal brain activity? What is the role of altered inhibition, excitation, or structure of brain network connections? What is the function and mechanism of the oscillatory interactions between key brain regions (especially during the sleep)?

Environmental modelling
From the many scales of atmospheric dynamics, the team mainly focuses on modelling local and microscale processes. In the area of inversion approaches in the **air quality modelling** we contributed to the development of the CMAQ-adjoint model, the first multiphase adjoint model of the regional scale chemical transport, as well as to its application for improved estimation of the spatial and temporal distribution of emissions and the responsibility of the emission sources for health risks. In the area of urban thermal comfort and air quality modelling we developed the first urban layer components for the well-known **microscale meteorological model PALM** and contributed to the development of the complete urban climate model PALM-4U. We work on ensemble filtering and covariance modelling, growing research areas in data assimilation dealing with high dimension, high demand of computational resources and model errors with sophisticated structures. A specific research area in environmental modelling concerns **local climate zones (LCZ)**. The concept of the LCZ emerged out of the need to improve the documentation of atmospheric heat island observations. Recent research pointed out that the influence of thermal, radiative, metabolic and surface-cover properties on the formation of local climate may significantly vary with respect to the geographical location of the zone, the size of the city, the position within the city and relief.

Main results in the evaluated period 2015-2019

The review is structured according to research directions; since the directions overlap (and naturally emerge from each other), there is some overlap between sections.

Nonlinear dynamics and causality
Any scientific discipline strives to explain the causes of observed phenomena. Methods based on Granger causality have been applied in diverse scientific fields from economics and finance, through Earth and climate sciences to brain research. The team has a long history in using information theory in order to generalize the Granger causality concept for nonlinear complex systems. It is always useful to have a comparison of performance of different approaches to causality detection using a set of benchmark data. In cooperation with the Slovak Academy of Sciences, causality detection methods based on information theory, dynamical systems theory, as well as the original Granger approach and its generalizations have been compared using various simulated systems (stochastic, deterministic, chaotic). Strengths and weaknesses of different methods have been described in [1].

Information transfer in real-world systems requires a finite time interval. In [2], the team compared two methods for estimating the information transfer delay in a numerical study using examples of coupled dynamical systems, demonstrating scenarios when the methods fail. This result is an important warning that the inference of coupling delay is a problem that is not yet fully solved and the results must be interpreted with utmost care, especially when making critical decisions.

The ultimate goal of the causality analysis is the distinction between the cause and the effect. In cooperation with the Slovak Academy of Sciences the team has discovered that the chaotic systems violate the causality principle that causes precede their effects. In computer simulations we reversed the direction of time in the analysed time series. In the case of linear systems, the mathematical methods for the detection of causality swapped the cause and the effect, as expected. However, in chaotic systems, the cause remains to be the cause also in the reversed direction of time and even if the cause occurred after the effect. The relation between time asymmetry and entropy production seems to explain this paradox, observed in computer simulations. If one observes this paradox in the causal analysis of real data, it may be a signal that the studied data resulted from a chaotic process. This finding helps uncover the physical mechanisms of causal relations observed in data. For instance, the causal influence of El Niño on weather in distant places can be characterized as a linear transfer of time-delayed signals, or, in physical language, a transport of air masses and related energy. On the other hand, in a human body, the rhythm of breathing influences the rhythm of the heart and this causal relation can be understood as interactions of nonlinear, possibly chaotic oscillators [3].

An important aspect of complex systems is that the dynamics and interactions occur at a multitude of spatial and/or temporal scales. This inspired Milan Paluš (Phys. Rev. Lett. 112 (7) 078702 (2014)) to propose a new approach to causality analysis: While traditionally, causal relations are sought between variables, now one can focus on a single time series reflecting an evolution of a multiscale process and infer causal relations between dynamical phenomena evolving on different time scales or frequencies (periods). The approach has found applications in climate science. Proposed original cross-scale causality analysis uncovered a causal relation between the phase of slow irregular oscillations with the period about 7-8 years, and the amplitude of faster fluctuations on time scales from a few months to 2-3 years in variability of the air temperature in European mid-latitudes. As a consequence, the amplitude of the annual air temperature cycle changes by 1˚C, while the changes in the air temperature annual means reach 1.5 ˚C. The effect is seasonally dependent with the maximum change observed in the boreal winter season. The difference between the winter mean temperatures in various parts of the 7-8-year cycle can exceed 4 ˚C. The large inter-annual changes can strongly influence economy, power engineering, agriculture, tourism and human lives in general [4]. Similarly, El Niño - Southern Oscillation is a gigantic natural pendulum which swings sea surface temperatures in the tropical Pacific from its cold (La Niña) to warm phase (El Niño). The amplitude of these swings also fluctuates. The extremes have a tremendous influence on the global climate, including droughts in Asia and Australia, and floods in South America. The period changes from two to seven years, providing an ultimate application field for our cross-scale causality analysis. With researchers from the University of Wisconsin-Milwaukee and the University of California San Diego the team has identified oscillatory components of El Niño, uncovered the causal network and showed how synchronization leads to extreme El Niño or La Niña events. This result [5] helps to better understand and potentially predict extreme El Niño events and their disastrous consequences. The concept and methods of the cross-scale causality has been reviewed by Milan Paluš [6], and it was also a topic of bilateral cooperation with the Chinese Academy of Sciences. In [7], an indication was obtained of the causal influence of the North Atlantic Oscillation on air temperature in central Europe.

Besides the information-theoretic approach, many other attempts rooted in nonlinear dynamics have tackled the inherent limitations of the original Granger concept (based on linear stochastic processes). Together with researchers from the University of Oldenburg, Jaroslav Hlinka introduced a rigorous extension of Granger causality to nonlinear systems by local linearization, redefining Granger causality as a function of position in the state space while providing a global index consistent with the original definition. This formal treatment bridges the key gap between linear and nonlinear causality estimates and avoids the problems of heuristic approaches. Its properties have been shown in a series of papers extending the initial result [8].

[1] Krakovská, A., Jakubík, J., Chvosteková, M., Coufal, David, Jajcay, Nikola, Paluš, Milan. Comparison of Six Methods for the Detection of Causality in a Bivariate Time Series. Physical Review E. 2018, 97(4), 042207. ISSN 2470-0045. Available: doi: 10.1103/PhysRevE.97.042207
[2] Coufal, David, Jakubík, J., Jajcay, Nikola, Hlinka, Jaroslav, Krakovská, A., Paluš, Milan. Detection of Coupling Delay: A Problem not yet Solved. Chaos. 2017, 27(8), 083109. ISSN 1054-1500. Available: doi: 10.1063/1.4997757
[3] Paluš, Milan, Krakovská, A., Jakubík, J., Chvosteková, M. Causality, Dynamical Systems and the Arrow of Time. Chaos. 2018, 28(7), 075307. ISSN 1054-1500. Available: doi: 10.1063/1.5019944
[4] Jajcay, Nikola, Hlinka, Jaroslav, Kravtsov, S., Tsonis, A.A., Paluš, Milan. Time Scales of the European Surface Air Temperature Variability: The Role of the 7-8 Year Cycle. Geophysical Research Letters. 2016, 43(2), 902-909. ISSN 0094-8276. Available: doi: 10.1002/2015GL067325
[5] Jajcay, Nikola, Kravtsov, S., Sugihara, G., Tsonis, A.A., Paluš, Milan. Synchronization and causality across time scales in El Niño Southern Oscillation. npj Climate and Atmospheric Science. 2018, 1(12 November), 33. E-ISSN 2397-3722. Available: doi: 10.1038/s41612-018-0043-7
[6] Paluš, Milan. Coupling in complex systems as information transfer across time scales. Philosophical Transactions of the Royal Society A-Mathematical Physical and Engineering Sciences. 2019, 377(2160), 20190094. ISSN 1364-503X. Available: doi: 10.1098/rsta.2019.0094
[7] Wang, G., Zhang, N., Fan, K., Paluš, M. Central European air temperature: driving force analysis and causal influence of NAO. Theoretical and Applied Climatology. 2019, 137(1-2), 1421-1427. ISSN 0177-798X. Available: doi: 10.1007/s00704-018-2676-1
[8] Wahl, B., Feudel, U., Hlinka, Jaroslav, Wächter, M., Peinke, J., Freund, J.A. Granger-Causality Maps of Diffusion Processes. Physical Review E. 2016, 93(2), 022213. ISSN 2470-0045. Available: doi: 10.1103/PhysRevE.93.022213

Network inference and analysis

Inferring the interactions within a pair of subsystems is in practice only a first step in the study of a system conceptualized as a network of interacting units. The study of real-world complex systems increasingly involves the study of their topological structure using graph theory. Here, two vertices (subsystems) are connected by an edge if the related time series are mutually dependent (undirected graphs) or one time series causally influences the other (oriented graphs). It is a well-known fact that many real-world systems have the small-world property: high local clustering and short paths between nodes. Members of the team has previously showed that correlation-based graph representations have a bias towards small-world structure and extended this to show that this small-world bias accounts for the small-world structure in both brain and climate data, quantitatively assessing a critical error repeated in this much publicized area of research into complex networks [9].

This problematic interpretability of the correlation-based networks is one of the motivations for resolving the full causal (directed) structure of interactions. A novel approach to quantification of causality in multivariate time series has been introduced and applied to global air pressure field in another study the team has published as a result of long-term collaboration with the Potsdam Institute for Climate Impact Research [10]. The method distinguishes and quantifies direct and mediated causal links and allows one to construct and to interpret a directed network representation.

The study, published in Nature Communications, already collected over 120 citations. The approach is suitable also for networks of size approximately up to a hundred nodes. Notably, this study utilized our earlier in-house result - a principled way to extract non-random correlation structures and carry out dimensionality reduction in multivariate climate data by means of spectral clustering [11]. For large networks, the use of heuristic assumptions allows effective estimation of causal networks, as was demonstrated in another study [12]. It allowed for an estimation of causal interactions of the global climate network, based on temperature in 2512 positions on the globe. A tailored approach facilitated the discovery of a smooth causal flow in the global climate network, that until then went unnoticed due to the inability of general-purpose causal network analysis methods to deal with such a large network.

The problem of estimation of network structure is further complicated by the potential nonlinear character of the coupling, rendering the standard linear methods invalid. However, they often still provide a very good proxy for the dependence pattern. The team previously discussed this issue in brain activity dynamics and climate dynamics, and recently extended this methodology to disentangle the sources and topological consequences of apparent nonlinearity in stock network data [13]. For substantially nonlinear systems, linear methods are not suitable; for such scenarios, Milan Paluš [14] proposes mutual information rate (MIR) as a measure of similarity of dynamics, extracting physically reasonable network structure from meteorological data.

Moreover, in case of large deviations from linearity, the structure of dependencies might not be any more well captured by pairwise interactions. In particular, redundant or synergic higher-order interactions might take place. Appropriately describing systems is a rapidly developing topic of research. in collaboration with researchers from the University of Calgary the team introduced a novel, efficient, and general method to quantify how good the assumption of pairwise interactions is, and provided this assumption holds, to reliably infer the backbone of the *direct* network connectivity between the interacting units. The method entails a novel entropy maximization scheme based on conditioning on entropies and mutual information, and is particularly suited for the typical case of large nonlinear systems where the dynamical units have more than two states and in the undersampled regime, both situations where previous methods failed. The advances were documented in a series of papers starting with initial introduction using phase oscillator networks and a resting-state human brain network as examples [15].

The team has also been involved in the study of complex nonstationary dynamics by demonstrating that an approach recently applied by the prominent group of Prof. Sporns to uncover **brain microstates** detects false states in pure white noise. This highlights critical methodological issues affecting a whole field of intensive research [16]. The major goal of analysing complex systems is supported by **graph theoretical research**. One example is the contribution to a long open question of classification of HH-homogeneous graphs given by [17]. The results represent symmetries based on extending a local morphism of their structure into a global one. The topic has been motivated by studies of homomorphisms by Lovasz when defining graph limits as limiting models for complex networks, and was later extended to study of symmetries of vertex colored graphs as those are models for multilayered complex networks. Considering graph structure, situations are explored when corresponding data structures, such as various matrices, are defined with uncertainties. Examples include studies of the radius of regularity, i.e. the distance of a matrix to a nearest singular one using Chebyshev norm. See [18] for the newest results concerning the application of an approximation algorithm based on semidefinite relaxation to several specific classes of matrices.

The team has also recently developed a new direction of applied research in data science, in particular concerning applications in **cybersecurity**. It has contributed towards the development of a novel method for identification of mislabeled samples via phantom nodes in label propagation [19], which has been patented in the USA and applied in production systems of the AVAST antivirus company. As part of the same collaboration, two more technologies have been developed. A novel method of indexing sandbox logs for mapping program behavior has been also applied and already received a US patent [20], while the last patent of the last technology is pending, concerning the application of convolutional malware classifiers to learn from raw executables. The collaboration with AVAST is continuing and recently we opened collaboration in area of data mining for cybersecurity with CISCO.

[9] Hlinka, Jaroslav, Hartman, David, Jajcay, Nikola, Tomeček, David, Tintěra, J., Paluš, Milan. Small-world Bias of Correlation Networks: From Brain to Climate. Chaos. 2017, 27(3), 035812. ISSN 1054-1500. Available: doi: 10.1063/1.4977951
[10] Runge, J., Petoukhov, V., Donges, J.F., Hlinka, Jaroslav, Jajcay, Nikola, Vejmelka, Martin, Hartman, David, Marwan, N., Paluš, Milan, Kurths, J. Identifying Causal Gateways and Mediators in Complex Spatio-Temporal Systems. Nature Communications. 2015, 6(7 October), 8502. ISSN 2041-1723.
Available: doi: 10.1038/ncomms9502
[11] Vejmelka, Martin, Pokorná, Lucie, Hlinka, Jaroslav, Hartman, David, Jajcay, Nikola, Paluš, Milan. Non-Random Correlation Structures and Dimensionality Reduction in Multivariate Climate Data. Climate Dynamics. 2015, 44(9-10), 2663-2682. ISSN 0930-7575. Available: doi: 10.1007/s00382-014-2244-z
[12] Hlinka, Jaroslav, Jajcay, Nikola, Hartman, David, Paluš, Milan. Smooth Information Flow in Temperature Climate Network Reflects Mass Transport. Chaos. 2017, 27(3), 035811. ISSN 1054-1500. Available: doi: 10.1063/1.4978028
[13] Hartman, David, Hlinka, Jaroslav. Nonlinearity in Stock Networks. Chaos. 2018, 28(8), 083127. ISSN 1054-1500. Available: doi: 10.1063/1.5023309
[14] Paluš, Milan. Linked by Dynamics: Wavelet-Based Mutual Information Rate as a Connectivity Measure and Scale-Specific Networks. In: Tsonis, A.A., ed. Advances in Nonlinear Geosciences. Cham: Springer, 2018, 427-463. Aegean Conferences. ISBN 978-3-319-58894-0.
Available: doi: 10.1007/978-3-319-58895-7_21
[15] Martin, E.A., Hlinka, Jaroslav, Davidsen, J. Pairwise Network Information and Nonlinear Correlations. Physical Review E. 2016, 94(4), 040301. ISSN 2470-0045. Available: doi: 10.1103/PhysRevE.94.040301
[16] Hlinka, Jaroslav, Hadrava, Michal. On the Danger of Detecting Network States in White Noise. Frontiers in Computational Neuroscience. 2015, 9(12 February), 11. ISSN 1662-5188. Available: doi: 10.3389/fncom.2015.00011
[17] Aranda, A., Hartman, David. The independence number of HH-homogeneous graphs and a classification of MB-homogeneous graphs. European Journal of Combinatorics. Online 29 November 2019, 103063. ISSN 0195-6698.
Available: doi: 10.1016/j.ejc.2019.103063
[18] Hartman, David, Hladík, M. Regularity Radius: Properties, Approximation and a Not A Priori Exponential Algorithm. Electronic Journal of Linear Algebra. 2018, 33(December), 122-136. ISSN 1537-9582.
Available: doi: 10.13001/1081-3810.3749

[19] Vejmelka, Martin. *Identification of mislabeled samples via phantom nodes in label propagation*. Avast Software, s. r. o. - Ústav informatiky AV ČR, v. v. i. 2019. Patent no.: US10198576. Date of the patent acceptance: 05.02.2019.
Available: http://hdl.handle.net/11104/0272996

[20] Vejmelka, Martin. *Distance and method of indexing sandbox logs for mapping program behavior*. Avast Software, s. r. o. - Ústav informatiky AV ČR, v. v. i. 2019. Patent no.: US10437986. Date of the patent acceptance: 08.10.2019.
Available: http://hdl.handle.net/11104/0272995

Neuroimaging data analysis and modelling

The analysis of brain imaging data poses a rich set of methodological challenges targeted by the team. One of them is the estimation of neuronal activity from observed indirect and noisy measurements by functional magnetic resonance imaging. The team developed a novel deconvolution method and demonstrated its performance in comparison with standard approaches on both simulated and real neuroimaging data [21], including a publicly available MATLAB Toolbox. Within a long-term collaboration with Prof. Mantini in KU Leuven, Jaroslav Hlinka contributed to the development of novel algorithms for artefact removal from simultaneously acquired EEG/fMRI measurements [22]. The team also regularly contributes to neuroscientific studies, including design, data analysis, result interpretation and methodological supervision. For instance, in a study of cannabis-induced altered states of consciousness [23] by construction of a fit-for-purpose permutation tests allowing one to detect significant differences in the brain state repertoire, and its relation to the drug levels in blood as well as subjectively reported experience; in another study by supervision of application and interpretation of the tract-based spatial statistics showing global, distributed brain changes already in the early stages of schizophrenia, and it proposed and implemented simulation pipeline explaining the appearance of localized results in previous studies with smaller samples [24]. The long list of published results from interdisciplinary neuroscience collaborations includes machine learning classification in **schizophrenia**, brain damage risk markers in **psychosis**, the effect of **wearing smart glasses** on the brain, the effect of physiotherapy on brain plasticity in **multiple sclerosis,** the prediction of psychotic episodes from **telemedicine** records, or the standardization of **depression** scales. Another area of research is the field of **computational neuroscience**, in particular modelling emergent phenomena in complex brain dynamics. One of the central questions concerns spatiotemporal brain activity patterns, in particular how does the pattern of functional connectivity (dependencies between remote neurophysiological events) emerge from the interplay of structural connectivity (brain anatomy) and local dynamics? The team collaborates with the group of Prof. Schöll at the Technical University of Berlin in understanding the **partial synchronization and chimera states** in brain networks under different approximations of realistic brain connectivity, see e.g. [25]. The is also a long-term collaboration with Prof. Sejnowski and Prof. Bazhenov in the area of **biophysically realistic simulations** of biological neural networks. Their recent paper tried to fill the gap in modelling of MEG/EEG measurements by 1) explicitly modelling spindle dynamics in a large network of neurons during sleep, 2) creating realistic whole-brain connectivity based on structural MRI and 3) describing the biophysical propagation of electromagnetic field following anatomical constraints, ultimately projected to signals measured by EEG/MEG on the scalp. They found that simulated sleep spindles were similar both in amplitude and topography to empirical measurements and thus such a multiscale model provides a framework on how to understand large scale integrative brain oscillations in terms of underlying neuronal activity [26]. Finally, the team is involved in international collaborative research into the **dynamics of epileptic seizures**. For example, recent work published in Nature Neuroscience [27] shows that transition to seizure is not necessarily a sudden phenomenon, but rather a slow process involving progressive loss of neuronal network resilience, governed by the ubiquitous principle of critical slowing prior to transition. Elucidating its governing principles by mathematical models and combined human, animal and in vitro study allowed us to unify previously antagonistic theories and observations about seizure genesis. The key input was in proposing an epileptic dynamics model capturing both the critical slowing down and the dual effect of perturbation on seizure initiation; the team also contributed to research design, data analysis and interpretation.

[21] Pidnebesna, Anna, Tomeček, David, Hlinka, Jaroslav. BRAD: Software for BRain Activity Detection from hemodynamic response. Computer Methods and Programs in Biomedicine. 2018, 156(March), 113-119. ISSN 0169-2607.
Available: doi: 10.1016/j.cmpb.2017.12.021
[22] Marino, M., Liu, Q., Koudelka, V., Porcaro, C., Hlinka, Jaroslav, Wenderoth, N., Mantini, D. Adaptive Optimal Basis Set for BCG Artifact Removal in Simultaneous EEG-fMRI. Scientific Reports. 2018, 8(June), 8902. ISSN 2045-2322.
Available: doi: 10.1038/s41598-018-27187-6
[23] Zaytseva, Y., Horáček, J., Hlinka, Jaroslav, Fajnerová, I., Androvičová, R., Tintěra, J., Salvi, V., Balíková, M., Hložek, T., Španiel, F., Páleníček, T. Cannabis-induced altered states of consciousness are associated with specific dynamic brain connectivity states. Journal of Psychopharmacology. 2019, 33(7), 811-821. ISSN 0269-8811. Available: doi: 10.1177/0269881119849814
[24] Melicher, T., Horáček, J., Hlinka, Jaroslav, Španiel, F., Tintěra, J., Ibrahim, I., Mikoláš, P., Novák, T., Mohr, P., Höschl, C. White Matter Changes in First Episode Psychosis and their Relation to the Size of Sample Studied: A DTI Study. Schizophrenia Research. 2015, 162(1-3), 22-28. ISSN 0920-9964.
Available: doi: 10.1016/j.schres.2015.01.029
[25] Chouzouris, T., Omelchenko, I., Zakharova, A., Hlinka, Jaroslav, Jiruška, Přemysl, Schöll, E. Chimera States in Brain Networks: Empirical Neural vs. Modular Fractal Connectivity. Chaos. 2018, 28(4), 045112. ISSN 1054-1500.
Available: doi: 10.1063/1.5009812
[26] Rosen, B. Q., Krishnan, G. P., Šanda, Pavel, Komarov, M., Sejnowski, T., Rulkov, N., Ulbert, I., Eross, L., Madsen, J., Devinsky, O., Doyle, W., Fabo, D., Cash, S., Bazhenov, M., Halgren, E. Simulating Human Sleep Spindle MEG and EEG from Ion Channel and Circuit Level Dynamics. Journal of Neuroscience Methods. 2019, 316(15 March), 46-57. ISSN 0165-0270.
Available: doi: 10.1016/j.jneumeth.2018.10.002
[27] Chang, W.C., Kudláček, Jan, Hlinka, Jaroslav, Chvojka, Jan, Hadrava, Michal, Kumpošt, Vojtěch, Powell, A. D., Janča, R., Maturana, M. I., Karoly, P. J., Freestone, D. R., Cook, M. J., Paluš, Milan, Otáhal, Jakub, Jefferys, J. G. R., Jiruška, Přemysl. Loss of neuronal network resilience precedes seizures and determines the ictogenic nature of interictal synaptic perturbations. Nature Neuroscience. 2018, 21(12), 1742-1752. ISSN 1097-6256.
Available: doi: 10.1038/s41593-018-0278-y

Environmental modelling

The main results include the development of the crucial components for simulating urban canopy in the LES model PALM: the Urban Surface Model, the multi-reflective 3-D Radiative Transfer Model and some other modelling components (plant canopy, air quality, biometeorology): [28], [29]. The part of the work also involved a thorough assessment and validation of the new developed model components. All these achievements were part of our strong international collaboration (e.g. Univ. Hannover, Berlin, Helsinki, KIT). The newly developed model has already been used for urban thermal comfort and air quality studies in Prague in cooperation with the Prague Municipality and the Prague Institute of Development and Planning. All mentioned papers which were submitted to the GMD journal (two published and four in public review) are the culmination of a years-long effort.

A long term multi-centric collaboration continues using the CMAQ-adjoint model with multiple results following the initial success of predicting premature deaths from emission propagation [30], with the plan to use them for emission assessment studies in CR in the near future. Successful collaboration with the University of Colorado Denver resulted in new theoretical results in data assimilation and ensemble filtering. The most significant are several convergence results, especially the proof of convergence of the square root ensemble Kalman filter [31], being the first to prove the convergence of this type of filter to the exact filtering distribution in the limit for the size of the ensemble going to infinity. This filter became popular in practice and the paper increases confidence in the proper statistical foundation of this technique. Several other results in the area dealt with the Ensemble Kalman smoother including convergence results, the relation of the smoother to standard variational method 4DVAR in data assimilation and development, asymptotic variances of the maximum likelihood estimates in hierarchical covariance models and an experimental validation of data assimilation method which approximates the covariance matrix using its spectral properties, including a proof that this method improves data assimilation when the method has a special structure. Notably, these results, besides being a contribution to modelling, have a practical impact, such as for use in the operational wildland fire warning system in Colorado [32], semiparametric statistical modelling of natural gas consumption adopted by market regulating authorities [33] or big data analysis of global navigation satellite data [34].

A specific area of research in environmental modelling concerns the topic of the **local climate zones**. One of the negative consequences of climate change in Central Europe is a more frequent occurrence of extreme weather, esp. heat waves [35], [36]. Temperatures in urban areas, where most of the population live, are additionally increased by so-called Urban Heat Islands (UHI). Municipalities are therefore compelled to look for tools, which can reduce the impact of temperature extremes (heat stress) on the population living in the cities [37]. Thermodynamic modelling in combination with spatial analysis in GIS can provide a very useful tool for municipalities. So-called “what-if” scenarios can provide essential and important information for municipalities, regional governments and other stakeholders on a local scale with consideration for a whole city (region). Modelling can easily identify and describe localities vulnerable to heat stress under projected future climates [35], socio-economic and demographic scenarios. This research was focused on identifying and categorizing localities vulnerable to high temperatures and heat stress in Central European Cities [37], [35], [36]. For these purposes we analysed air temperatures and heat stress indices simulations in the manner of standardized approach of local climate zones (LCZ) concept. These analyses were followed by analyses of temperatures and heat stress indices variability in particular LCZ's and carried out for different Representative Concentration Pathway scenarios, urban development scenarios and demographic projections [35].

[28] Resler, Jaroslav, Krč, Pavel, Belda, Michal, Juruš, Pavel, Benešová, N., Lopata, J., Vlček, O., Damašková, D., Eben, Kryštof, Derbek, P., Maronga, P., Kanani-Sühring, F. PALM-USM v1.0: A New Urban Surface Model Integrated into the PALM Large-eddy Simulation Model. Geoscientific Model Development. 2017, 10(10), 3635-3659. ISSN 1991-959X.
Available: doi: 10.5194/gmd-10-3635-2017
[29] Maronga, B., Banzhaf, S., Burmeister, C., Fuka, V., Geletič, Jan, Krč, Pavel, Resler, Jaroslav. Overview of the PALM Model System 6.0. Geoscientific Model Development. 2020, 13(3), 1335-1372. ISSN 1991-959X.
Available: doi: 10.5194/gmd-13-1335-2020
[30] Turner M.D., Henze, D.K., Capps, S., Hakami, A., Zhao, S., Resler, Jaroslav, Carmichael, G., Stanier, C., Baek, J., Sandu, A., Russell, A.G., Nenes, A., Pinder, R., Napelenok, S., Bash, J., Percell, P., Chai, T. Premature Deaths Attributed to Source Specific BC Emissions in Six Urban US Regions. Environmental Research Letters. 2015, 10(11), 114014. ISSN 1748-9326.
Available: doi: 10.1088/1748-9326/10/11/114014
[31] Kwiatkowski, E., Mandel, Jan. Convergence of the Square Root Ensemble Kalman Filter in the Large Ensemble Limit. SIAM/ASA Journal on Uncertainty Quantification. 2015, 3(1), 1-17. ISSN 2166-2525.
Available: doi: 10.1137/140965363
[32] Vejmelka, Martin, Kochanski, A., Mandel, Jan. Data Assimilation of Dead Fuel Moisture Observations from Remote automated Weather Stations. International Journal of Wildland Fire. 2016, 25(5), 558-568. ISSN 1049-8001.
Available: doi: 10.1071/WF14085
[33] Konár, Ondřej, Malý, Marek, Brabec, Marek, Pelikán, Emil, Kasanický, Ivan. Spolupráce na vyhlášce o pravidlech trhu s plynem č. 349/2015 Sb. 2015 (Cooperation on the Public Notice on the Natural Gas Market Regulations No. 349/2015).
Available: http://www.eru.cz/cs/-/vyhlaska-c-349-2015-sb-o-pravidlech-trhu-s-plynem
[34] Douša, J., Eliáš, M., Václavovič, P., Eben, Kryštof, Krč, Pavel. A Two-stage Tropospheric Correction Model Combining Data from GNSS and Numerical Weather Model. GPS Solutions. 2018, 22(3), 77. ISSN 1080-5370.
Available: doi: 10.1007/s10291-018-0742-x
[35] Geletič, Jan, Lehnert, M., Dobrovolný, Petr, Žuvela-Aloise, M. Spatial modelling of summer climate indices based on local climate zones: expected changes in the future climate of Brno, Czech Republic. Climatic Change. 2019, 152(3-4), 487-502. ISSN 0165-0009. Available: doi: 10.1007/s10584-018-2353-5
[36] Geletič, Jan, Lehnert, M., Savić, S., Milošević, D. Inter-/intra-zonal seasonal variability of the surface urban heat island based on local climate zones in three central European cities. Building and Environment. 2019, 156(June), 21-32. ISSN 0360-1323. Available: doi: 10.1016/j.buildenv.2019.04.011
[37] Geletič, Jan, Lehnert, M., Savić, S., Milošević, D. Modelled spatiotemporal variability of outdoor thermal comfort in local climate zones of the city of Brno, Czech Republic. Science of the Total Environment. 2018, 624(15 May), 385-395. ISSN 0048-9697. Available: doi: 10.1016/j.scitotenv.2017.12.076