DEPARTMENT OF MATHEMATICS

SELF-ASSESSMENT RESEARCH

2015-2020

Summary & Case Studies

VU VRIJE UNIVERSITEIT AMSTERDAM

SUMMARY

The Department of Mathematics at the Vrije Universiteit Amsterdam is part of the Faculty of Science. Due to increasing student numbers and investments through the *sectorplan Beta,* the department has grown substantially over the past years, and currently employs 35 assistant, associate and full professors. The department has no formal subdivision into groups.

The central strength of the department, and core to its strategy, is the intertwining of fundamentals and applications. The academic culture is such that the entire spectrum is valued and appreciated, from theorems on abstract algebraic structures to optimized algorithms for patient distribution over hospitals. Both research and teaching in the department cover this broad spectrum, with a dual aim to have impact in academic research and valorisation in society. Apart from our research objectives, we also strive for an excellent curriculum for our BSc and MSc undergraduates; a strong teaching environment and a strong research environment go hand in hand.

The excellent research in dynamical systems and in stochastics (probability, statistics, analytics & optimization) form the two main pillars of our department, accompanied by thriving activity in topology, geometry and algebra. The department advances mathematics and its applications in the broadest sense. Research on fundamental mathematical problems goes together with developing novel mathematical models, methods and algorithms to tackle questions in business and finance, in logistics, in health care, in neuroscience, in biology, in medicine, in physics, in engineering, in forensics and even in legal issues. The department actively reaches out to societal partners to engage in research collaborations. For example, there is a long line of research on the planning and scheduling of emergency services including ambulances, and statistical methodology generated by our research is being employed in the Amsterdam UMC. Other examples are our work on the energy transition and its relation to power networks, elderly care, and suicide prevention.

The strategy of the department is grounded in the core values of the Vrije Universiteit Amsterdam: responsible, open and personally engaged. The culture in the department is to work in a collegial fashion. The division of labour is primarily based on consultation and consent. Where possible we take advantage of individual talents. As a result, the hierarchy in the department is relatively flat, with most responsibilities related to day-to-day work lying with the staff members themselves. We give full freedom to our staff members to choose their research projects. In our view, this is the best guarantee for excellent research as well as societal impact. We also stimulate a collaborative academic culture, where collaborations can be within the department, with mathematicians all around the world, with researchers in other fields, as well as with partners in industry and other organizations. We do this by showing appreciation for this full range of activities, both on a daily basis and in criteria for promotions, and by providing the necessary financial support, including travel funds, partial matching of investments in PhD positions, etc. The policy of the department is to have all papers published under open access, either through one of the deals between large academic publishers and Dutch universities, or by publishing both the preprint and the author accepted version on the arXiv and/or the VU Research Portal repository.

Looking ahead, the department has a solid organization and financial basis to build on. With an influx of excellent young professors, the department has a bright future. It will be vital to guide the next generation into leadership roles, with a special eye towards (gender) diversity. The “Recognition and Rewards” policy recently adopted by the university allows for personalized development paths. While there is clearly work to do to put this into practice at the department level and to make the opportunities and boundaries concrete, the direction of this dynamics aligns very well with the long-term strategy of our department.

Case Study 1
TOPOLOGICAL DATA ANALYSIS

Topological data analysis (TDA) is a branch of data science which applies topology to study the *shape* of data, i.e., the coarse-scale, global, non-linear geometric features of data. Examples of such features include clusters, loops, and tendrils in point cloud data, as well as modes and ridges in functional data. While the history of TDA dates back to the 1990s, in recent years the field has advanced rapidly, leading to a rich theoretical foundation, highly efficient algorithms and software, and many applications.

To get an idea of how topology can be applied to data, consider the discrete set of data points shown in Figure 1a. There is clearly a 'circular structure' to the data and our pattern-seeking brains have no trouble inferring such information. Topology per se cannot infer this circularity (the topology of a finite set is rather uninteresting), so one proceeds by thickening every point by a certain thickness as shown in Figure 1e. The resulting annulus has a well-defined circular structure that can be measured by means of *homology vector spaces*, a classical tool from algebraic topology. However, one must be careful to choose a radius within the right range: if the radius is too small then the space is not fully connected

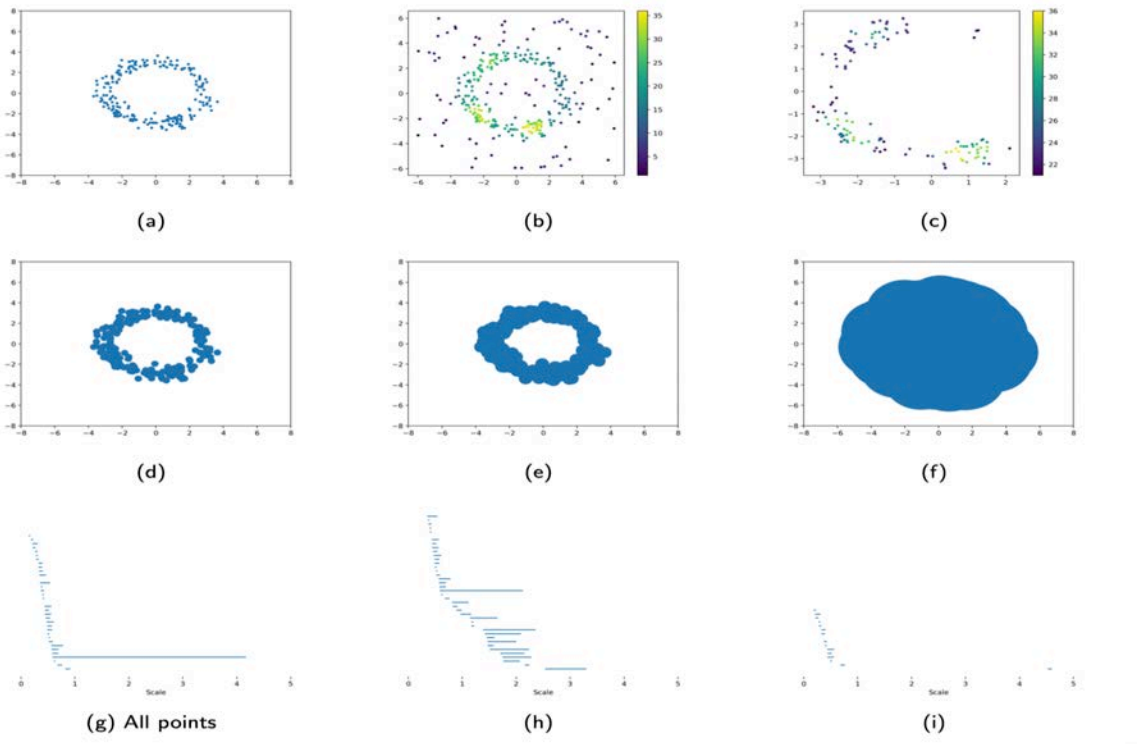

Figure 1. *Illustration of TDA. The points in* ***(b)*** *and* ***(c)*** *are colored by density.*

(Figure 1d), and if it is too large then the hole has vanished altogether (Figure 1f). For more complicated data, it is not at all clear how one should choose the right scale, or if such a scale even exists. Therefore *persistent* homology was introduced in order to track the topological features across scales, similar to how a dendrogram is used in hierarchical clustering methods. The output of persistent homology is a *barcode.* The barcodes of the point clouds in Figures 1a-c are shown in Figures 1g-i. An interval starting at scale b that ends at scale d, is to be interpreted as a feature that appears when the points are thickened by a radius $b/2$ and that vanishes when the radius reaches $d/2$. We see that the long interval in Figure 1g vanishes around $d = 4$, accurately reflecting the fact that the points seem to lie on an annulus of inner diameter 4. While the barcode is *stable* with respect to perturbation of the data, it is extremely sensitive to outliers. Consider the same points as before but with a few additional points scattered around (Figure 1b) and the associated barcode (Figure 1h); the circular structure can no longer be deduced from the barcode. A potential way to rectify this problem is to consider only points above a certain density threshold, but this would in turn be very sensitive to the choice of the threshold. See for instance Figure 1c and its associated barcode in Figure 1i. Ideally one would thus have a tool that allows one to deduce topological signatures across both scale and density in a *parameter-free* way,

and that is precisely what the field of *multi-parameter* persistent homology is about. In the multi-parameter setting, one filters a dataset not only by scale but also by one or more additional parameters. In addition to the example given above, there are many settings for which such constructions appear naturally, like time-varying data, data with tendrils, and functional data (e.g., image data).

The major challenge in this more general setting is that a significant increase in algebraic complexity precludes the existence of a barcode, which in turn has severely limited the development of tools for the data scientist. The good news is that multi-parameter persistent homology is a very active field of research, with many invariants introduced in recent years, and one can find yet more in the classical literature on commutative algebra and representation theory. The main question for TDA is thus which of those invariants can be useful in the development of data analysis methodology, i.e. invariants that are both descriptive and efficiently computable.

Impact

While the research at the VU concerns multiple aspects of TDA, the core focus lies with the development of novel invariants and distances in multi-parameter persistent homology, as well as their connection to the field of representation theory of algebras. We now mention a few contributions that have been well-received by the community.

1) Botnan (VU), Oppermann (Trondheim) and Oudot (Paris) recently have shown that there exists a stable barcode-analogue for multi-filtrations. While its interpretation is less intuitive, it is expected that many of the tools used to perform inference on barcodes in the one-parameter setting can be made to work in this setting too.
2) Botnan and Crawley-Boevey (Bielefeld) proved a fundamental structure theorem for persistent homology indexed by any partially ordered set.
3) Bauer (Munich), Botnan, Oppermann and Steen (Santa Cruz) gave a complete description of the representation theory of certain algebraic objects inspired by data analysis. This result is particularly interesting as it follows from a novel result about *cotorsion torsion triples* in abelian categories; a good example of how problems in data science inspire deep results in pure mathematics, and in this case, in the fields representation theory of algebras and homological algebra .
4) In TDA, the *interleaving distance* has received a significant amount of attention, as it is, in a certain sense, the optimal stable distance. Work by Botnan, Kerber (Graz) and Bjerkevik (Trondheim/Graz) shows that it is NP-hard to approximate the interleaving distance within a factor of 3.

There is a growing interest in TDA in the Netherlands, and for that reason, Botnan and Barthel (VU) are organizing *a day of applied topology* at the VU. This event will bring together researchers and practitioners of applied topology in the Netherlands. In July 2021, Botnan, Bauer, and Lesnick (Albany) organized a workshop on multi-parameter persistent homology at the Lorentz Center. Additionally, Botnan, Hirsch (Groningen/Aarhus), and Bonnet (Groningen) organized a workshop on the interplay between topology and stochastics in Groningen in September 2021.

Botnan gave a lecture series on TDA at ICRA 2020 and at the GQT summer school of 2021, in addition to numerous (keynote) talks at international workshops and conferences. An interview about Botnan and his research on topological data analysis appeared in the May 2021 edition of the magazine *Forskerforum* (the monthly publication of The Norwegian Association of Researchers).

Selected publications

[1] U. Bauer, M.B. Botnan and S. Oppermann, J. Steen (2020). Cotorsion torsion triples and the representation theory of filtered hierarchical clustering. *Advances in Mathematics* 369, 107-171.
[2] H.B. Bjerkevik, M.B. Botnan and M. Kerber (2019). Computing the interleaving distance is NP-hard. *Foundations of computational mathematics*, 1-35.
[3] M.B. Botnan and W. Crawley-Boevey (2020). Decomposition of persistence modules. *Proceedings of the American Mathematical Society* 148.11, 4581-4596.
[4] M.B. Botnan, S. Oppermann and S. Oudot. (2021). Signed barcodes for multi-parameter persistence via rank decompositions and rank-exact resolutions. *arXiv preprint* arXiv:2107.06800.

Case Study 2
COMPUTER ASSISTED PROOFS IN DYNAMICS

Introduction

In dynamical systems all the action is due to nonlinearity. It is the source of the enormous variety of phenomena displayed by these system, while the other side of the medal is that the nonlinear character usually makes pen-and-paper analysis incredibly difficult. With modern-day scientific computing algorithms and hardware, one can almost 'see' the solutions. But since discretization approximations are inevitable in simulations, it is not obvious at all that what you see is what is true.

The field of rigorous numerics aims to bridge this gap by supplementing numerical calculations with mathematically guaranteed, explicit bounds on the discrepancy between truth and approximation. This leads to computer-assisted proofs and thus to theorems which fit into the fabric of mathematics: such results can be relied upon to build mathematical theory. Computer-assisted proofs (CAPs) are by now a widely used tool in mathematics, with the famous proof of the four-color theorem and the resolution of Kepler's densest sphere packing problem as prime examples. In the theory of dynamical systems the role of computer-assisted proofs goes back to the proof of the Feigenbaum conjecture, and later the existence of the strange attractor in the Lorenz system.

Idea

Applicability of the general theory of dynamical systems to concrete nonlinear models often hinges on obtaining detailed information on special solutions. The last decade has seen a surge in the use of computer-assisted proofs to obtain this information. Indeed, adopting a general functional analytic point of view, one starts by reformulating the problem of locating the solution of interest as a zero finding problem in some suitably chosen Banach space. One then proves that an associated Newton-like fixed point operator contracts in a small neigborhood of an approximate solution, obtained from a numerical simulation. There are many variants of this approach, but invariably these involve a combination of two aspects. First, some large but finite computation is done on a truncated problem, with rounding errors controlled via interval arithmetic. Second, a careful pen-and-paper analysis of the truncation error is performed, controlling the infinite-dimensional 'tail' of the problem. Together these then give an explicit estimate on the distance between the true solution and its numerical approximate counterpart.

At the VU Department of Mathematics, working with a variety of colleagues, Jan Bouwe van den Berg has developed and refined this approach to obtain results in nonlinear dynamics described by ordinary, partial and delay differential equations [1]. This was supported by an NWO Vici grant and has contributed to the success of four PhD trajectories.

Connecting orbit problems in ordinary differential equations (ODEs)

Connecting orbits describe transitions between different states of a system. They serve as relatively simple building blocks for complicated dynamics or spatial patterns, as they can often be combined into more complex structures. But constructing the building blocks is not an easy task. Even in cases where major parts of a problem can be analyzed using, for example, asymptotic analysis, the crucial final step of finding the connections often requires computer-assistance. Three examples where Van den Berg and collaborators obtained a CAP to finish such an analysis are a connecting orbit that describes target patterns in the Swift-Hohenberg problem [2], a connecting orbit that describes a spiral wave in the complex Ginzburg-Landau equation (based on [6]), and a connecting orbit that describes domain walls between hexagonal spots and stripes [3] in a generalized Swift-Hohenberg equation with broken symmetry (Figure 1).

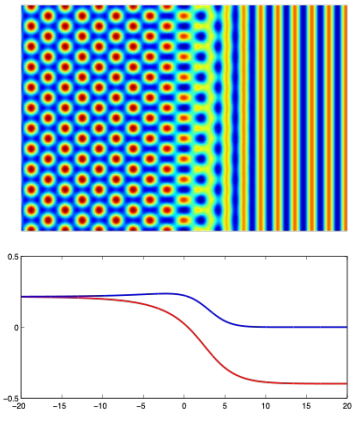

Figure 1. *A transition between a hexagonal spot pattern and a stripe pattern in a modified Swift-Hohenberg equation. The connecting orbit in the system of ODEs which results from an asymptotic analysis, was found using a CAP and is depicted as the pair of red and blue graphs.*

Bifurcations in delay-differential equations (DDEs)

In situations where the response of a system is not instantaneous, time delays need to be incorporated in the models, leading to DDEs. Such problems can be posed in a functional analytic setting in much the same way as ODEs, although the required analysis is more subtle. One recent result concerns the resolution of a long-standing conjecture in delay equations, called Wright's conjecture. Partial progress had already been made by many researchers in the field of DDEs. Through a computer-assisted quantitative analysis of the curve of slowly oscillating solutions emanating from a Hopf bifurcation point, Van den Berg and Jaquette were able close the final gap and turn the conjecture into a theorem [4].

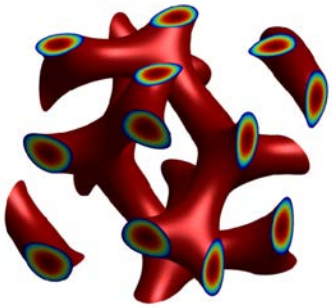

Figure 2. *A so-called "double gyroid" pattern in the Ohta-Kawasaki model for di-block copolymers. It is a highly-symmetric configuration which is found to be stable in certain parameter regimes.*

Periodic patterns in partial differential equations (PDEs)

Since the pivot of the CAP approach is a zero finding problem, it is equally suitable for problems with more than one spatial or time-like variable, such as PDEs. Of course the estimates, algorithms and computational cost are then more involved. Periodic patterns with intricate symmetries in the Ohta-Kawasaki model for di-block copolymers have been proven by Van den Berg and Williams using these computer-assisted method [5], and their stability has also been assessed (Figure 2). Analogously, time-periodic solutions of the forced Navier-Stokes equations for incompressible fluid flows were established in [7], see Figure 3.

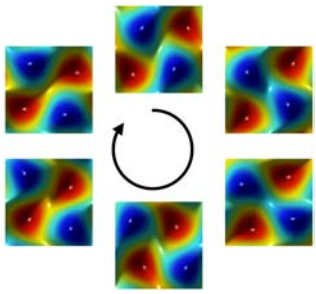

Figure 3. *Snapshots of the vorticity of a time-periodic solution of the Navier-Stokes equations with Taylor-Green forcing and periodic boundary conditions in two spatial dimensions.*

Impact

Fixed point formulations for studying problems in nonlinear analysis are ubiquitous. For computer-assisted proofs this approach goes back half a century. Nevertheless, based on increased computational power and improved algorithms and analysis, the last decade has seen a flurry of exciting computer-assisted results in nonlinear dynamics. Jan Bouwe van den Berg has made a substantial contribution to this development together with a diverse group of international collaborators: Jean-Philippe Lessard (McGill, Montreal), J.F. Williams (SFU, Vancouver), Jay Mireles James (FAU, Florida), Jonathan Jaquette (Boston) and Maxime Breden (Ecole polytechnique, Paris). These results have made the dynamical systems community aware that turning numerical simulation into rigorous mathematical results is a feasible proposition: not trivial, but accessible on a laptop without the need for highly specialised code. Such CAPs lead to results that are interesting in their own right (e.g., when analysing solutions to "normal form" equations), but they also fill gaps in already existing analysis. Moreover, they generate mathematical guarantees for solutions of dynamically different types than those reachable via pen-and-paper methods such as closed-form formulas. These can then, for example, be used as benchmarks for new numerical methods.

Van den Berg has organized a series of collaborative international workshops at Leiden, Banff (Canada) and Palo Alto (USA), which generated a lot of research activity, provided networking opportunities for junior mathematicians, and contributed to community building. The impact of the work of Van den Berg is also illustrated by the invitation to organize an AMS short course at the 2016 annual meeting of the American Mathematical Society, as well as a plenary lecture at *Equadiff* 2019.

Selected publications

[1] J.B. van den Berg and J.-P. Lessard (2015). Rigorous numerics in dynamics. *Notices of the American Mathematical Society* 62, 1057-1061.
[2] J.B. van den Berg, C.M. Groothedde and J.F. Williams (2015). Rigorous computation of a radially symmetric localized solution in a Ginzburg-Landau problem. *SIAM Journal on Applied Dynamical Systems* 14, 423-447.
[3] J.B. van den Berg, A. Deschênes, J.-P. Lessard and J.D. Mireles James (2015). Stationary coexistence of hexagons and rolls via rigorous computations. *SIAM Journal on Applied Dynamical Systems* 14, 942-979.
[4] J.B. van den Berg and J. Jaquette (2018). A proof of Wright's conjecture. *Journal of Differential Equations* 264, 7412-7464.
[5] J.B. van den Berg and J.F. Williams (2019). Rigorously computing symmetric stationary states of the Ohta-Kawasaki problem in three dimensions. *SIAM Journal on Mathematical Analysis* 51, 131-158.
[6] J.B. van den Berg, W. Hetebrij and B. Rink (2020), The parametrization method for center manifolds. *Journal of Differential Equations* 269, 2132-2184.
[7] J.B. van den Berg, M. Breden, J.-P. Lessard and L. van Veen (2021). Spontaneous periodic orbits in the Navier-Stokes flow, *Journal of Nonlinear Science* 31:41, 1-64.

Case Study 3
CONLEY THEORY AND THE ALGEBRA OF DYNAMICAL SYSTEMS

Introduction

Dynamical systems are mostly defined via systems of differential equations, or as iterations of continuous maps. In practice these systems are too complex to study without any computational tools. A first insight into the workings of a dynamical system is often provided by numerical simulation in order to understand, for example, the long term behaviour of the system, which has been extremely successful in the application of dynamical systems as a prediction tool. Such analysis does not give much insight into global behaviour, however, nor does it provide rigorous results about the dynamical structure of a system. Computational Conley theory makes various aspects of global dynamics computable using algebra, combinatorics and algebraic topology.

Scientific problem

In the past decades, discretization methods have been developed in which combinatorialization and algebraic topology are the key ingredients, cf. [1,8,9,11]. The first steps in this direction have proved very promising. We mention in particular the early work in [1]. Motivated by the combinatorial/algebraic topological approach, a program has been started to study dynamics from the point of order theory and universal algebra, cf. [3,4,5]. The goal is to find a formulation of dynamics which provides a natural formalism to study finite renderings of dynamics. The latter should be interpreted as combinatorializations of dynamics, providing rigorous information of the full system within a given resolution. One key ingredient of dynamics that is pivotal in the algebraic approach, is the dichotomy between gradient-like and recurrent dynamics. For example, consider the motion of a natural pendulum. As the pendulum swings, it cannot return to its original position with the same velocity because it loses energy due to friction—gradient-like dynamics. In the absence of friction the system returns to its initial state, and its motion is periodic. In this case the system is said to exhibit recurrence. Understanding this dichotomy between recurrent and nonrecurrent or gradient-like behaviour in a system is central to the study of dynamical systems.

Mathematical approach

The work of Rob van der Vorst and collaborators in [2,3,4,5,6] is focused on treating dynamical systems as algebraic structures. A pivotal ingredient is the attractor of a dynamical system. An attractor is a subset of phase space which ‘attracts’ nearby dynamics. A system has many (or infinitely many) attrac-

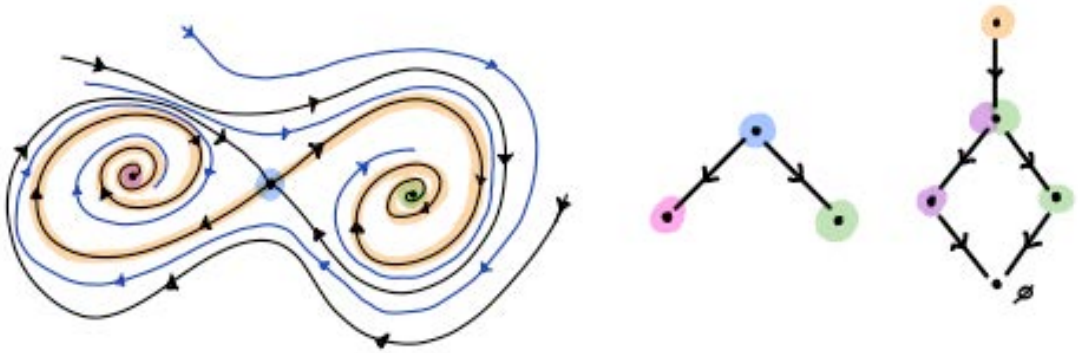

Figure 1. *A flow that displays a number of different attractors. The green and purple fixed points are attractors that locally attract neighboring points. The blue middle saddle point together with the connecting orbits (beige) gives the global attractor of the system [left]. This is an example of a global decomposition of dynamics [middle]. The algebraic structure of the attractors is presented by the lattice of attractors given as a Hasse diagram [right].*

tors in general. The set of all attractors forms a bounded, distributive lattice which is a key algebraic structure. Via lattice theory one can understand the decomposition of dynamics into smaller irreducible parts, but more importantly, it opens the door to an algorithmic analysis of dynamics as an order structure that can be studied via its finite substructures. Finding a (finite) lattice of attractors is equivalent to a decomposition of dynamics within a certain resolution, cf. Figure 1. The latter provides an algebraic

method for discretizing dynamical systems. Figure 2 below shows a discretization of a two-dimensional population model. The lattice of all attractors of the discretized model is a sublattice of the lattice of attractors of the system. The decomposition obtained from the discretized model represents an actual decomposition of the full system. Our current research is concerned with further strengthening and extending the mathematical foundation of the algebraic approach towards dynamics. Some of the subjects are:

- Boolean algebras with operators: this algebraic structure play a pivotal role in modal logic and turns out to be the fundamental tool to reformulate dynamics. Boolean algebras are very useful for formulating 'discretization' and 'combinatorialization' in algebraic terms, cf. [7].
- Sheaf cohomology: the algebraic structures such as the lattice of attractors of a dynamical system change as the parameters in the system change. One can reformulate this by regarding the space of dynamical systems as a parameter space and to study the local robustness of the lattice of attractors. This leads to the formulation of the sheaf of attractors which describes the attractor lattices as a 'function over the dynamical systems'. An important invariant for sheaves is sheaf cohomology. In [6] we link sheaf cohomology to bifurcations.
- Sheaf theory and discretization: we can combine sheaf theory with the discretization approach via Boolean algebras. The goal is to construct a 'finite' sheaf theory such as cellular sheaves which allow us to extend the computational techniques to include parameter dependence.

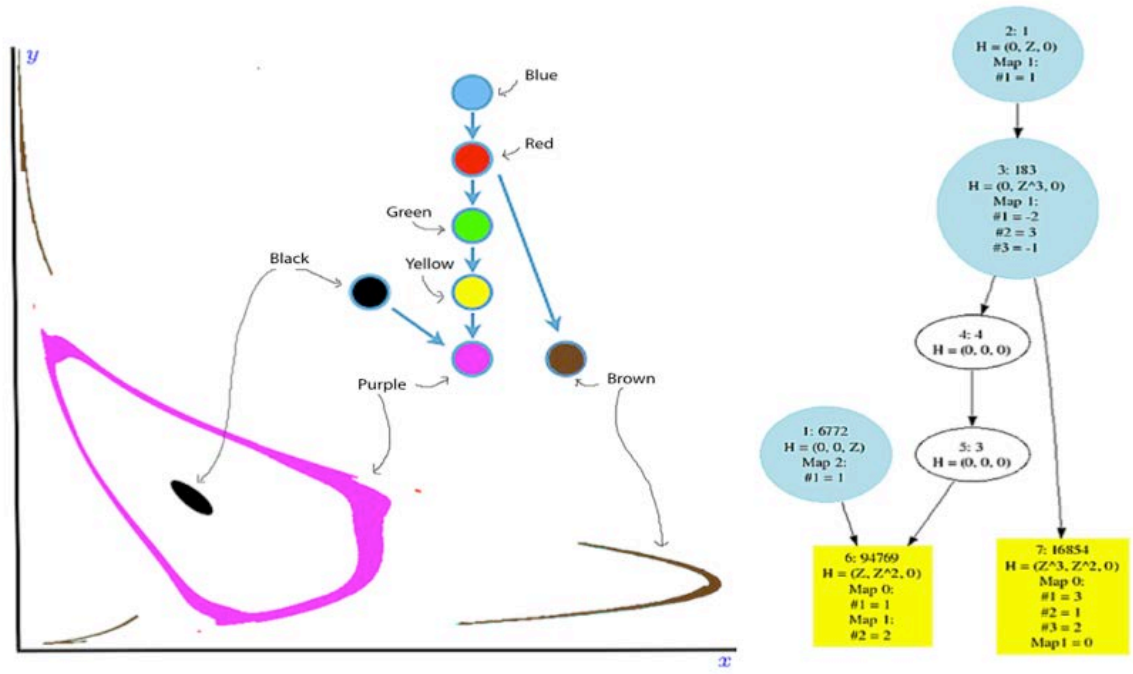

Figure 2. A discretization of a two-dimensional population model. *The middle graph provides an order representation of the global dynamics. The right picture combines order with algebraic topological data (Conley indices) which reveals local information about the 'nodes' in the graph. The nodes represent invariant sets for the dynamics and the arrows display the direction of the dynamics (gradient-like part). We distinguish attracting, repelling and saddle-like behaviour in the above decomposition.*

Collaborations

Over the past years Van der Vorst has collaborated extensively with W.D. Kalies (Florida Atlantic University) and K. Mischaikow (Rutger's University). For the period 2020-2025 prof. Mischaikow has a visiting position at our department. He will visit us yearly for a one month period combined with two prolonged visits of one semester. The research concerning Boolean algebras as reformulation of dynamics is joint work with K. Spenlove (Oxford University). With K. Spendlove we also study an important

topological invariant for dynamics, the Conley index. In this setting we are mainly interested in the representation of dynamics in terms of so-called 'connection matrices'. There is a relation between the latter and generalizations of spectral sequences. The subject of sheaf cohomology is joint work with K.A. Dowling (Rutgers University) and W.D. Kalies. This research now focuses on merging sheaf theory with discretization.

Selected publications

[1] Z. Arai, W.D. Kalies, H. Kokubu, K. Mischaikow, H. Oka, P. Pilarczyk (2009). A database schema for the analysis of global dynamics of multiparameter systems. *SIAM Journal on Applied Dynamical Systems*, 8(3), 757-789.
[2] W.D. Kalies, K. Mischaikow, and R.C.A.M. Vandervorst (2005). An algorithmic approach to chain recurrence. *Found. Comput. Math.*, 5(4), 409-449.
[3] W.D. Kalies, K. Mischaikow, and R.C.A.M. Vandervorst (2014). Lattice structures for attractors I. *J. Comput. Dyn.*, 1(2), 307-338.
[4] W.D. Kalies, K. Mischaikow, and R.C.A.M. Vandervorst (2015). Lattice structures for attractors II. *Found. Comput. Math.*, 1(2), 1-41.
[5] W.D. Kalies, K. Mischaikow, and R.C.A.M. Vandervorst (2021). Lattice structures for attractors III. *J. Dynam. Differential Equations*, https://doi.org/10.1007/s10884-021-10056-8
[6] K.A. Dowling, W.D. Kalies, K. Mischaikow, and R.C.A.M. Vandervorst, Continuation sheaves in dynamics: sheaf cohomology and bifurcation, ArXiv: 2106.8478.
[7] K. Spendlove and R.C.A.M. Vandervorst, Closure algebra discretizations, flow topologies and parabolic dynamics, in preparation.
[8] E. Boczko, W.D. Kalies, and K. Mischaikow (2007). Polygonal approximation of flows. *Topology Appl.*, 154(13), 501-520.
[9] M. Mrozek (1999). An algorithm approach to the Conley index theory. *J. Dynam. Differential Equations*, 11(4), 711-734.
[10] T. Kaczynski, K. Mischaikow, and M. Mrozek (2004). Computational homology, volume 157 of *Applied Mathematical Sciences*. Springer-Verlag, New York.
[11] S. Day (2003). A rigorous numerical method in infinite dimensions. *Ph.D. dissertation*.

Case Study 4
DISTRIBUTED STATISTICAL METHODS FOR HIGH-DIMENSIONAL MODELS

Distributed methods

Both in statistics and machine learning there has been substantial interest in the design and study of distributed statistical or learning methods in recent years. One driving reason is the fact that in certain applications datasets have become so large that it is often unfeasible, or computationally undesirable, to carry out the analysis on a single machine. In a distributed method the data are divided over a cluster consisting of several machines and/or cores. The machines in the cluster then process their data locally, after which the local results are somehow aggregated on a central machine to finally produce the overall outcome of the statistical analysis, cf. Figure 1.

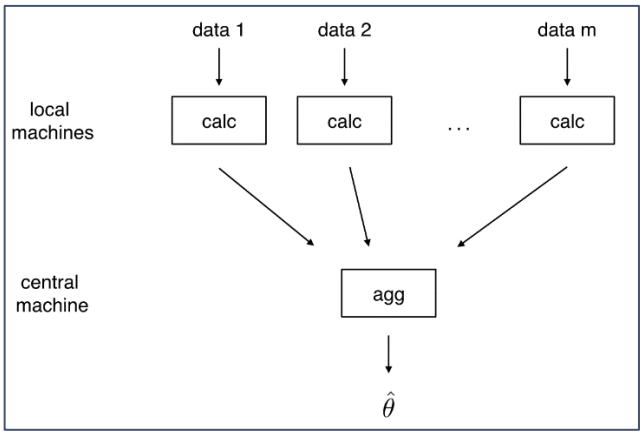

Figure 1. *Distributed parameter estimation.*

Distributed methods are not only used for computational reasons, but are for instance also of interest in situations where privacy is important and it is undesirable that all data are handled at a single location. Moreover, there are applications in which data are by construction collected at multiple locations and first processed locally, before being combined at a central location.

Suboptimal performance of naive procedures

It is not difficult to see that in distributed settings where the number of local machines is relatively large, it is suboptimal to naively use a statistically optimal procedure in the local machines and then average the local results in the central machines. This typically results in underfitting and misleading uncertainty quantification (UQ). Figure 2 illustrates this in a simple signal reconstruction problem. The black line is the true signal, which is observed in white noise. In the right-hand panel an optimal (Bayesian) signal reconstruction is shown (solid blue line), together with uncertainty quantification (blue dashed lines). On the left the same data was distributed over 20 local machines, the same optimal procedure was used in every local machine and then the result was averaged in the central machine. Clearly, the resulting signal reconstruction (solid red line) is much worse and the UQ (red dashed lines) is way off.

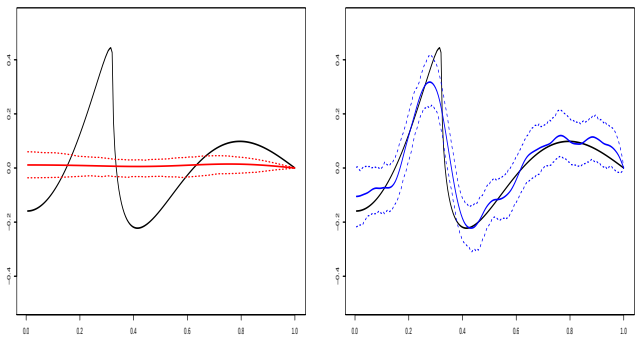

Figure 2. *Performance of naive distributed vs non-distributed method.*

Performance of existing methods

As the example illustrates, distributed statistical methods for high-dimensional problems should be designed and tuned differently than classical, non-distributed methods. Until very recently, there was little or no mathematical understanding of this phenomenon. In particular, it was not clear whether distributed methods that have been proposed over the years actually succeed in avoiding the indicated problems. In a paper in *Journal of Machine Learning Theory* Van Zanten and Szabo (presently at Bocconi) studied a number of existing distributed Bayesian methods, including widely used methods such as *Consensus Monte Carlo, WASP,* and *distributed GP's*. They proved rigorous mathematical statements asserting that there is substantial difference between the performance of the various methods. Figure 3 illustrates the results for 6 different methods. Some give signal reconstruction and uncertainty quantification comparable to the optimal non-distributed method, but some do significantly worse in one or both respects.

Fundamental possibilities and limitations

The analysis of existing distributed methods raises the question: what are the fundamental statistical limits of distributed methods? For classical methods there exist well-developed statistical theory that explains in many situations how the best convergence rate that a statistical procedure can attain is related to the complexity of the object that is being learned. In their recent paper in the *Annals of Statistics* Van Zanten and Szabo initiated the development of such a theory for distributed methods. In particular, they studied how the best possible performance of methods is related to the amount of communication that is allowed between the local and central machines. A first interesting finding is that there are different regimes, depending on the relation between the total sample size (or signal-to-noise ratio) *n*, the number of local machines *m*, the amount of bits *B* of allowed communication, and the complexity of the unknown object that is being learned. In the signal reconstruction problem this complexity is measured by the smoothness *s* of the signal. In the paper it is proved that in this case, if *B* > *n^(1/(1+2s)),* then the amount of communication is large enough for distributed methods to attain the same rate as optimal non-distributed methods. If, on the other extreme, *B* < (n/m^(2+2s))^(1/(1+2s)), then the allowed communication is so little that the (slow) rates obtained by local machines cannot be improved by aggregating their results in the central machine. The third, intermediate, regime is in some sense the most interesting. In that range the communication limitation does deteriorate the best possible rate, but combining the results of the local machines does help. In the paper, Van Zanten and Szabo exhibit a concrete distributed procedure that attains the optimal rate in this case.

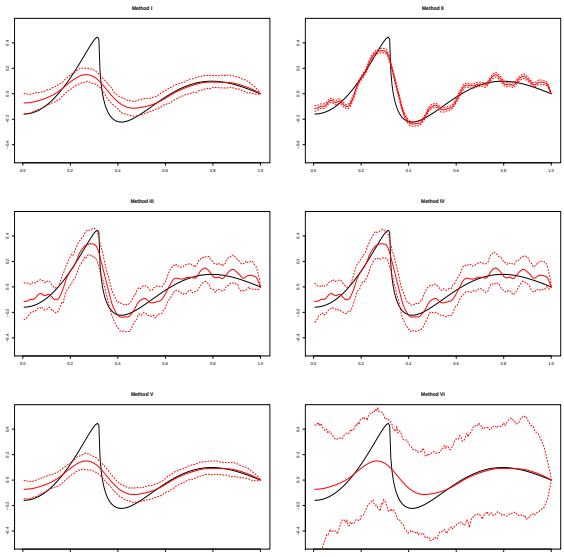

Figure 3. *Performance of various distributed methods.*

Adaptation to the complexity of the unknown structure

As with all statistical procedures in high dimensions, distributed methods have to be tuned carefully to achieve an optimal bias-variance trade-off, or, in other words, to avoid over- and underfitting. A key question is whether this can be done without knowing crucial complexity information, like the smoothness *s* in the signal reconstruction problem. It is by now well known that the answer is affirmative in the non-distributed setting, and there exist several methods to construct so-called *adaptive* procedures that do not use this information, including wavelet thresholding, Lepski's method, and hierarchical or empirical Bayes procedures. In distributed settings with limited communication the matter is, however, much more delicate, since the local machines cannot always extract enough information about the complexity from their smaller, local datasets. We have shown nonetheless that in the signal reconstruction problem, for instance, adaptation is possible over a non-trivial range of smoothness levels that depends on *n, m* and *B*. The procedure that we have constructed is not yet very practical. Designing practical, adaptive distributed methods is an important topic in ongoing research.

Another topic that is currently under investigation in a joint PhD project with Szabo and L. Vuursteen and A.W. van der Vaart (TU Delft), is the testing of statistical hypotheses in distributed settings. Here similar phenomena arise, but there are also interesting new aspects. For instance, it turns out that in a distributed framework signal detection, which is a particular testing problem, can benefit from the availability of a so-called *public coin*: a common source of randomness accessible to all local machines. We have provided a detailed explanation of how the detectability of a signal depends on a combination of the signal strength, the number of machines, the signal-to-noise ratio, the communication allowed and the availability of a public coin. A related current research direction is the application of these ideas to statistical meta-analysis, i.e., the combination of statistical testing outcomes (such as p-values) of independent studies.

Outlook: underpinning uncertainty quantification in distributed settings

Arguably, the most important topic in this area of research is the design of mathematically principled uncertainty quantification. Already in the non-distributed setting, it is known that this is extremely challenging if there is no auxiliary information about the complexity of the unknown structure. Distributed methods now routinely produce uncertainty quantification, but without any performance guarantees. With the increasing use of distributed methods in critical (e.g., medical) applications it is crucial that mathematical theory is developed that can fundamentally underpin such methods, and that can provide guidelines for their design and tuning.

Selected publications

[1] B.T. Szabo and J.H. van Zanten (2019). An asymptotic analysis of distributed nonparametric methods. *J. Mach. Learn. Res.* 20, 1-30.
[2] B.T. Szabo and J.H. van Zanten (2020). Adaptive distributed methods under communication constraints. *Ann. Statist.* 48 (4), 2347-2380.
[3] B.T. Szabo, L. Vuursteen and J.H. van Zanten (2021). Optimal distributed testing in high-dimensional Gaussian models. Under revision for *IEEE Transactions on Information Theory*.

Case Study 5

MAXIMIZING GROWTH RATE OF MICROORGANISMS

The importance of microbes

Single-celled organisms such as bacteria and yeasts, together referred to as microbes, are vital for humans. They are an essential part of human health and physiology, recycle oxygen, carbon and nitrogen on a global scale, and are essential in the production of medicines and foods such as yoghurt, beer, wine and soy sauce. They can of course also be a risk for human health. It is therefore important to be able to predict how microbes behave: in industrial applications, we wish to control which flavors or medicines are produced by microbes; growing microbes are susceptible to antibiotitics, but dormant cells can survive this and can become dangerous when they start growing again. For about thirty years, scientists have been trying to make such predictions bottom-up: by integrating detailed knowledge of the inner workings of cells, and extrapolating these to changes in cellular physiology and behaviour.

Collaboration with the Systems Biology group

For the last ten years, Bob Planqué and Joost Hulshof have been collaborating closely with the Systems Biology group at the VU, headed by Frank Bruggeman and Bas Teusink. After an initial period in which we studied the dynamical details of glycolysis in yeast, together with PhD student Gosse Overal, our attention has shifted towards whole-cell models. The evolutionary success of single-celled organisms is very closely tied to their cellular growth rate: high fitness means high metabolic rates, and high self-replication rates. We therefore try to understand how microbes can achieve steady states of fast growth, whether such states are uniquely defined by the environment, and how they can achieve optimality when changing from one environment to the next. Our theory is informed by experiments at the Systems Biology group, and also leads back to new experiments in their labs.

Using only stoichiometry

Traditional approaches to growth rate maximization have centered around the stoichiometric matrix of a microorganism. In such a matrix, each column represents one reaction, and each row one metabolite (small cellular compound, such as sugars, ATP, amino acids). A column in this matrix stipulates how much substrate and product is converted in one particular reaction event. The resulting optimization problem of finding states with high flux through the metabolic network then can be formulated as a pure Linear Programming problem. It can be shown that the set of vectors over which we wish to maximize is a convex pointed cone that is spanned by extreme rays. These rays are called Elementary Flux Modes (EFMs), minimal pathways connecting the beginning and end of the network. Hence, maximisers are EFMs. Even in this simple setting, new insights could be obtained. PhD student Daan de Groot, co-supervised by Planqué, showed that EFMs are indeed expressed by live microbes. Moreover, the change from pure respiration to a combination of respiration and fermentation (which happens when one gets muscle ache, occurs in cancer tumor cells, and also in yeasts producing alcohol when we brew beer) turned out only to depend on the number of constraints that are active, not on the identity of these constraints. This resolved a decades-old controversy in the literature, and has led to a new generation of experiments to determine which physico-chemical or biological constraints are actually hit by growing microbes.

Adding self-replication

The biggest challenge to develop a completely general theory for maximal growth rate lies in the fact that the growth rate of cells is an emergent property of the complete machinery with which cells produce new copies of themselves. In a long effort spanning four years, Planqué and Hulshof, together with De Groot, Bruggeman and Teusink, went completely back to square one and modelled the growth rate from first principles. This led to a new nonlinear optimization problem for growth rate μ, of the form

$$max\{\mu \mid A(\boldsymbol{c},\mu)\boldsymbol{\alpha} = \mathbf{0}, \boldsymbol{\mu} = \alpha_n g_n(\boldsymbol{c}), \boldsymbol{\alpha} \geq \mathbf{0}\}.$$

Identifying solutions to the constraint equations with the same support gives rise to a new type of elementary mode, which we termed Elementary Growth Modes (EGMs). These are the direct extensions of the more familiar EFMs, and growth rate maximizers are now EGMs, cf. Figure 1. This theory gives

the first general mathematical foundation for detailed computational models, and also has given us a way to describe the degrees of freedom cells have to self-replicate.

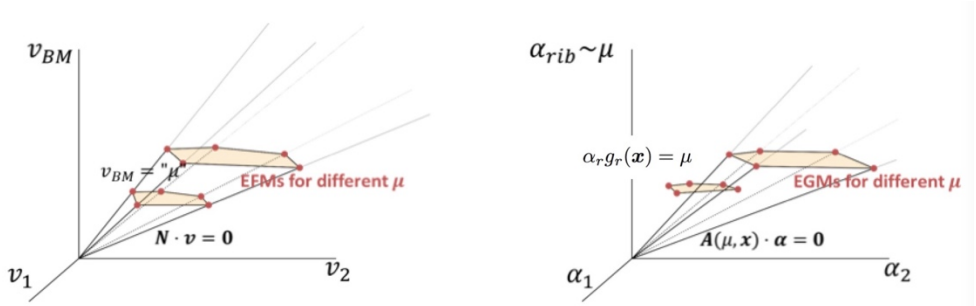

Figure 1. *Comparing EFM cones to EGM 'cones'. Left, a standard convex pointed cone spanned by extreme rays called EFMs. Right, a few instances of the solution space for different growth rates. Those with the same support need to be identified and form an EGM.*

Adaptive control

Cells that express only the enzymes involved in one of these elementary modes have already optimized their behaviour to a large extent. But within one of these modes, enzyme concentrations may still be chosen differently, leading to different steady state reaction rates or growth rates. It has become increasingly clear that living bacterial cells are able to actively tune these concentrations to maximize the growth rate. They are even able to regain such optimal states after changes in the environment, without directly sensing these changes. We questioned what type of control cells must implement biochemically in order to solve this problem.

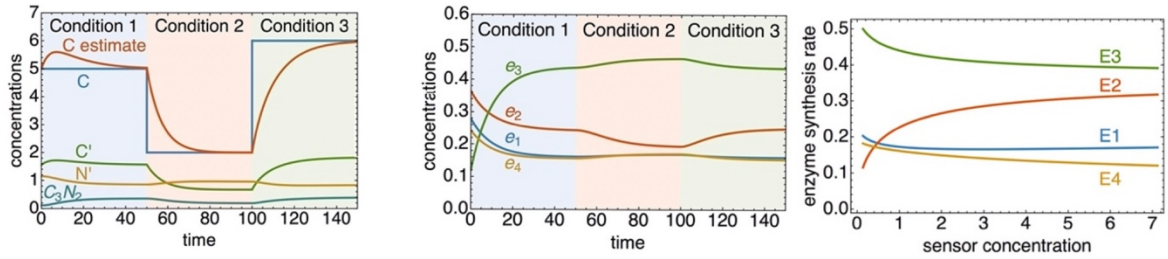

Figure 2. *Example dynamics of the adaptive control. External concentrations are varied from Condition 1 through 3. Enzyme concentrations change, causing the pathway to settle into a new (optimal) steady state. The enzyme synthesis rates, as functions of an internal 'sensor', are shown on the right, and are the same in all conditions.*

Together with MSc student Johan Hendriks, we developed a general framework for this. First, Planqué and Hulshof showed that this optimization problem can be reformulated as a convex problem. Dynamically, the concentrations of the metabolites $\boldsymbol{x}$ (small molecules that are produced and consumed by enzymatic reactions) change through an equation of the form $\boldsymbol{x}' = \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{e})$. Enzyme concentrations $\boldsymbol{e}$ change in a much simpler fashion, through synthesis and dilution by growth, and so their dynamical equations are given by $\boldsymbol{e}' = \boldsymbol{E} - \mu\boldsymbol{e}$ for some suitable function $\boldsymbol{E}$. If enzyme synthesis rates change according to some feedback from metabolism, then we need that $\boldsymbol{E} = \boldsymbol{E}(\boldsymbol{x})$. Remarkably, it is possible to construct for a given EFM with prescribed $\boldsymbol{F}(\boldsymbol{x}, \boldsymbol{e})$ a (*fixed, so environment-independent*) function $\boldsymbol{E}(\boldsymbol{x})$ with the property that the combined dynamical system for $\boldsymbol{x}$ and $\boldsymbol{e}$ has a unique steady state with optimal steady state flux. This is remarkable because the optimal state depends on the environment, but $\boldsymbol{E}(\boldsymbol{x})$ does not. The construction actually gives information about the number and identity of the metabolites that appear in $\boldsymbol{E}(\boldsymbol{x})$. Initial numerical experiments showed that for rather arbitrary EFMs this control seemed to be globally stable, see Figure 2 for an example. Planqué was recently awarded a grant to develop this work further, and together with Hulshof and PhD student Maarten Droste, he is currently trying to prove such global stability properties for individual EFMs.

Teaching

The collaboration with the Systems Biology group has also been very fruitful in our teaching. Planqué regularly co-supervises MSc thesis projects of Bruggeman, some of which have turned into publications. In addition, Planqué and Hulshof teach a yearly modelling workshop for third year BSc students on mathematical systems biology. The course is extremely rich in mathematical techniques—in four weeks students apply techniques from graph theory, linear algebra, convex analysis, dynamical systems, linear and nonlinear programming and control theory. It thus gives students a great experience in how the breadth of the mathematics they have just learnt can be applied fruitfully in just one context.

Selected publications

[1] G. B. Overal, B. Teusink, F. J. Bruggeman, J. Hulshof and R. Planqué (2018). Understanding startup problems in yeast glycolysis. *Mathematical Biosciences* 299, 117-126.
[2] M. T. Wortel, H. Peters, J. Hulshof and F. J. Bruggeman (2014) Metabolic states with maximal specific rate carry flux through an elementary flux mode. (2014). FEBS Journal 28, 1547-1555.
[3] R. Planqué, J. Hulshof, B. Teusink, J. C. Hendriks and F. J. Bruggeman (2018). Maintaining maximal metabolic flux by gene expression control. *PLoS Computational Biology* 14(9), e1006412.
[4] D. H. de Groot, C. van Boxtel, R. Planqué, F. J. Bruggeman and B. Teusink (2019). The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLoS Computational Biology* 15(3), e1006858.
[5] F. J. Bruggeman, R. Planqué, D. Molenaar and B. Teusink (2020). Searching for principles of microbial physiology. *FEMS Microbiology Reviews* 44(6), 821-844.
[6] D. H. de Groot, J. Lischke, R. Muolo, R. Planqué, F. J. Bruggeman and B. Teusink (2020). The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. *Cell. Mol. Life Sciences.* **77**, 441-453.
[7] D. H. de Groot, J. Hulshof, B. Teusink, F. J. Bruggeman and R. Planqué (2020). Elementary Growth Modes provide a molecular description of cellular self-replication. *PLoS Computational Biology* 16(1), e1007559.
[8] T. J. Clement, E. Baalhuis, F. J. Bruggeman, B. Teusink, R. Planqué and D. H. de Groot (2020). Unlocking Elementary Conversion Modes: ecmtool unveils all capabilities of metabolic networks. *Patterns.* 2(1), 100177.

Case Study 6
SAVING LIVES WITH MATHEMATICS: From reactive to proactive planning of emergency services

Societal problem

In serious life-threatening situations where every second counts, getting an ambulance to the scene on time or not can mean the difference between survival and death. The response-time requirement for high-emergency calls is country-specific, but is usually of the form "*In x% of the cases, the arrival time of an ambulance vehicle may not be longer than y minutes*". For example, in the Netherlands, x = 95 and y = 15. The problem is that in practice, such requirements are (too) often not met, with all its consequences for the patient-in-need. In order to realize short response times at an affordable cost, careful planning of ambulance services is necessary.

When realizing an efficient planning of the ambulance service, many questions come up. For example: How can we accurately predict the number of help requests depending on time and location? How can we properly anticipate and respond to peaks in demand for ambulance rides? How many base locations do we need, and what are the optimal locations? How many units (vehicles, drivers, medical personnel) should we deploy to achieve the desired level of service, and where and when? How can we do suitable personnel planning in such a way that all hard and soft preconditions are met? How can we achieve high coverage of ambulances across a region at any time of the day by using smart dynamic and pro-active real-time repositioning of ambulances?

Scientific problem

The above questions lead to a wide variety of scientific challenges. A highly complicating, but scientifically very challenging, factor in solving these kinds of planning issues is the *omnipresence of the phenomenon of uncertainty that is inherent in almost all facets of the ambulance service process*. The problem is that existing planning methods usually assume that the demand and availability of ambulances are known in advance, while these methods are usually very sensitive to changes in these input parameters. Underestimating these uncertain factors leads to inefficient, and therefore expensive, planning of ambulance rides. This requires groundbreaking research to develop new, scalable, forecasting and planning methods that are robust against changes in environmental factors.

Vision and solution concept

The key to reducing response times for the ambulance service is to make a **paradigm shift from the traditional reactive to a proactive** way of working. This so-called Dynamic Ambulance Management (DAM) entails two important decisions that ambulance service providers need to make:

1. *Dynamic dispatching*: "Which ambulance do we send to a given incident?", "Is it always optimal to send the closest-idle ambulance first, and if not, which ambulance should be sent?"
2. *Proactive relocation*: "How to proactively reposition ambulance vehicles to anticipate future incidents better?"

Solution approach

In the project *From Reactive Planning of Ambulance Services* (*REPRO)*, co-funded by NWO and five ambulance services providers in the greater Amsterdam area, we have developed stochastic optimiza-

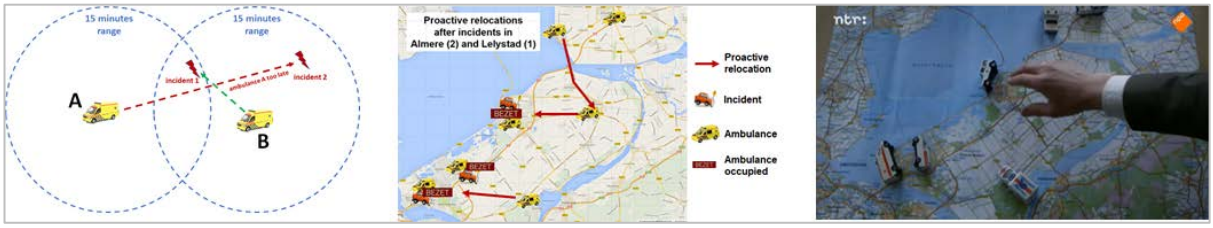

Figure 1. *Illustration of the analysis of dispatching (l) and proactive relocation (m) by means of toy examples (r).*

tion models for optimal decision making for dispatching and relocation of ambulances. Since the decision making needs to be fast, a key requirement is that the algorithms provide answers in real-time, not only in toy examples but also for realistic real-life scenarios. Figure 1 illustrates the dispatching and relocation problems and the way of working.

Mathematical models and publications

Our research involves a variety of models and solution techniques in the realm of forecasting methods, deterministic and stochastic optimization models. More specifically, we have developed machine learning models for forecasting incidents [8], facility location models (e.g., maximum expected coverage location models) for the strategic planning of base locations [3], and efficient heuristics for optimal dispatching [5, 6] and dynamic relocation of vehicles [1, 2, 7]. We emphasize that, in line with the nature of this research, the results have not only been published in classical Operations Research journals, but also in leading journals focused on the application of the results in real-life practice [3, 4].

Results and real-life pilots

Over the years, we have developed and evaluated a range of stochastic optimization algorithms for optimal dispatching and relocation of vehicles, as well as models for optimal locations of base stations. In doing so, the key to getting these algorithms implemented in real-life practice is that we have collaborated closely with our partnering ambulance providers. In particular, we have set up a real-life pilot in which the algorithms were tested (in collaboration with GGD Flevoland) in an iterative process of model improvement based on feedback from practice. This iterative process converged into effective algorithms that are well-aligned with the need and limitations of practice.

Implementation in practice and spin-off company

The model and methods have been implemented in a software package that is commercially exploited in the market, via the spin-off company Stokhos Emergency Mathematics. Currently, the algorithms are operational in several safety regions in the Netherlands.

Figure 2. *Illustration of the operational tools for relocation of ambulances (l) and firefighters (r).*

Other application areas

The work on ambulance planning has attracted the interest of other application areas as well, such as firefighter and police services. With the Fire Brigade Amsterdam/Amstelland, we have developed models for optimal locations of fire stations and for optimal relocation of firetrucks in case of large fire incidents, which has been the basis for a relocation tool named FireScore, currently deployed in the greater Amsterdam area. For police, we have developed forecasting methods for creating risk maps for high-impact crimes, such as burglaries, and robberies. For a railway infrastructure provider, we have developed models for optimally balancing preventive maintenance tasks and emergency response, as part of a public-private partnership.

PhD projects, academic staff involved and Huibregtsen Prize

This research (which was done partly in collaboration with CWI and TU Delft) has led to six PhD projects, out of which four have been successfully finished, and there are two more to come. The supervision of the PhD students was done by three full professors and a postdoctoral researcher.

Partly based on this line of research, profs. Sandjai Bhulai and Rob van der Mei were awarded the prestigious Huibregtsen Prize 2021, a national award for excellent research that is applied in real-life practice. The jury report states (paraphrase): “The mathematical solutions offered are efficient and elegant. What is truly exceptional within this field is how much effort Bhulai and Van der Mei put into ensuring that their work is used in practice. This is truly mathematics for the real world”.

Selected publications

[1] T.C. van Barneveld (2016). The minimum expected penalty relocation problem for the computation of compliance tables for ambulance vehicles. *INFORMS Journal on Computing* 28(2), 370-384.
[2] T.C. van Barneveld, S. Bhulai and R.D. van der Mei (2016). The effect of ambulance relocations on the performance of ambulance service providers. *European Journal of Operational Research* 252(1), 257–269.
[3] P.L. van den Berg, G.A.G. Legemaate and R.D. van der Mei (2017). Boosting the response-time performance of firefighter services by Mathematical Programming. *Interfaces* 47, 352-361.
[4] M. van Buuren, C.J. Jagtenberg, T.C. van Barneveld, R.D. van der Mei and S. Bhulai (2018). Ambulance dispatch center pilots proactive relocation policies to enhance effectiveness. *Interfaces* 48(3), 235-246.
[5] C.J. Jagtenberg, P.L. van den Berg and R.D. van der Mei (2017). Benchmarking online dispatch algorithms for Emergency Medical Services. *European Journal of Operational Research* 258, 715-725.
[6] C.J. Jagtenberg, S. Bhulai and R.D. van der Mei (2016). Dynamic ambulance dispatching: is the closest-idle policy always optimal? *Healthcare Management Science*, 1-15.
[7] C.J. Jagtenberg, S. Bhulai and R.D. van der Mei (2015). An efficient heuristic for real-time ambulance redeployment. *Operations Research in Health Care* 4, 27-35.
[8] G. Legemaate, J. de Deijn, S. Bhulai and R.D. van der Mei (2021). Severe weather-based fire department incident forecasting. To appear in *International Journal on Advances in Intelligent Systems*. A short version of this paper received the best-paper award at the *IARIA Data Analytics conference* (Porto, Portugal, 2019).