Success Formulas

Applications of Mathematics

Matthias Ehrhardt, Michael Günther, Wil Schilders

December 2021

""Mathematics is like oxygen. If it is there, you do not notice it.
If it would not be there, you realize that you cannot do without."

Contents

Preface

This book aims to show a little of the mathematics we encounter every day. It turns out that mathematics is present in sports, medicine and weather forecasting, traffic, and can even help solve murder cases. The book also shows that mathematics is important in business, where it often makes invisible contributions to visible successes. For example, mathematics contributes to innovation in all cutting-edge sectors, it is a so-called. key enabling technology In addition, nine prominent representatives will talk about the role of mathematics in their work and personal lives.

> *„We use mathematics every day. To predict the weather, measure time, conduct monetary transactions. Mathematics is more than just formulas and equations. It is reason. With it, our minds can solve the greatest puzzles we know.“*
>
> (off-text Charlie Eppes in TV series Numb3rs)

In this book, we want to demonstrate that mathematics is much more than pure calculation, simple insertion of values into formulas, solving puzzles, or discussing curves. Mathematics is present in many areas of our lives: it models and simulates the course of the Corona pandemic, it describes the effects and effectiveness of vaccination strategies, it controls robots, it optimizes production processes and traffic flows, and much more.

If you would like to know more about this book, please do not hesitate to contact us. On the related website https://erfolgsformeln.uni-wuppertal.de/ you will find an overview of the mathematicians who contributed to this book, as well as further literature on the individual chapters of this book. We have given the individual chapters the three levels of difficulty: easy '*',medium '**', difficult '***'.

The idea for this book arose from a brochure 'Succesformules' by Dutch mathematicians led by Wil Schilders, who was Mittelsten-Scheid Visiting Professor at the Bergische Universität Wuppertal in the winter semester 2020/21. We would like to take this opportunity to express our sincere thanks for the support of the Bergische Universität Wuppertal.

We hope you enjoy reading the chapters and interviews in this book.

Wuppertal,
February 2022

Matthias Ehrhardt
Michael Günther
Wil Schilders

1 Interview with Dr. Thomas Hahn

A brief summary of your career

Dr. Thomas Hahn has been Chief Expert Software at Siemens AG since 2011. After studying computer science at Friedrich Alexander University in Erlangen, he joined the company in 1986. Thomas Hahn has held various roles in manufacturing, such as head of software development for industrial automation systems, or in mobility, such as head of development for traffic control systems. In addition to his current position as Chief Expert Software, he was Head of Business Analytics and Monitoring from 2011 to August 2013.

In addition to his work at Siemens, Thomas Hahn is a member or board member of various committees, including Openlab CERN and the Plattform Industrie 4.0 steering committee. He is vice president of the OPC Foundation, vice chairman of GAIA-X AISBL, chairman of the Labs Network Industrie 4.0, board member of the Bavarian „KI-Rat“ and president of the Big Data Value Association.

What has mathematics meant to your career?

Mathematics as a basic science has been important to me in that it helps you think and reason in a structured and logical way. And these are important qualities to master in business, even if it doesn't involve mathematical relationships at all, such as business models, development processes or software architectures. You have to convince others of your own ideas or critically reflect on the ideas of others, and a systematic approach like the one you learned in mathematics simply helps enormously here.

What role does mathematics play in your institute/company? And more generally in the industry, in which you work?

At Siemens, mathematics plays different roles:Ḟirst, there are several mathematicians at Siemens who are absolute experts in their fields and are also recognized accordingly in the global community. They mostly act as drivers for this area from a certain application perspective.

There are also many engineers at Siemens who perform tasks that require a sound mathematical understanding, for example when it comes to developing specific simulation models or simulation calculations. Here, an understanding of both the tools used and the application context is then necessary in addition to a basic methodological understanding.

Finally, there are mathematicians who have ended their involvement with mathematics as such by joining the company, but who now use the methodological knowledge they acquired in their studies in practice.

Figure 1: Dr. Thomas Hahn

Are, in your view, mathematicians active enough to build bridges between university and business/society?

I would like to use this question to basically address the bridge between university and business. What I am saying applies not only to mathematics, but also to science and engineering. In business, it is extremely important to understand economic relationships. This applies not only to business or economics, but especially to the impact of engineering decisions on business. You can discuss technical solutions at length and in detail, but ultimately you will only be successful as a company with a technical solution if you can explain to a customer the benefits that the technical solution will bring him and what the unique selling point of the technical solution is compared to competitors. In my view, this knowledge is not taught enough at universities. However, I also see it as the duty of industry to become suitably involved in teaching at universities in order to convey such connections to students, preferably illustrated with concrete examples.

If you look at the ratio between interns or working students at our company and the number of students in total, it would be especially great if even more mathematicians would take the chance to gain practical experience already during their studies.

What advice do you have for young people regarding mathematics?

My first advice to young people is always to distinguish between „calculus“, which you learn in elementary school and the first grades of secondary school, and „mathematics“, which you don't really encounter until the last grades of school, but at the latest in college. For some, the first semester of math at university is simply a culture shock.

My second point is that you have to distinguish between mathematics as theory and mathematics as application in different domains. In business, it is predominantly the application of mathematics that is important. Application here often means the realization of mathematical algorithms in software, in this respect I can only recommend everyone to learn programming early and to gain as much experience as possible especially in this area.

And my last comment is that you don't always have to understand everything in mathematics either. I proofread a mathematics book myself as a student assistant, and I still didn't understand one proof in it, even though I went through the chain of reasoning several times with the professor at the time. Well, I found my way even without understanding this proof.

2 Interview with Dr. Markus Hoschek

A brief summary of your career

I studied mathematics and electrical engineering at the TU Darmstadt and ENSIMAG in Grenoble, France, and then worked as a research assistant at the TU Darmstadt and Siemens AG Munich until I received my PhD in 1999. Subsequently, I worked as a consultant for IBM international, advising banks on the development of risk models and the acquisition of companies.

Since 2008, I have been a member of the Executive Board of HEAG Holding AG - Beteiligungsmanagement der Wissenschaftsstadt Darmstadt (HEAG). HEAG sees itself as a multi-service provider and is, on the one hand, the holding and management company of the HEAG Group, headquartered in Darmstadt, with total assets of around €2,917 million and around 3,000 employees in 2019. In addition, it is the investment management company of the City of Science Darmstadt. In this context, HEAG supports Darmstadt's commercial holdings and anchors the city's objectives, especially in the more than 100 majority holdings with around 8,000 employees. HEAG is thus the central and authoritative advisor to the Science City of Darmstadt in all economic matters relating to the city's economy.

In connection with my work on the Executive Board of HEAG, I am a member of numerous supervisory bodies and advisory boards as well as associations of the municipal economy. I have also been Chairman of the Board of the Darmstadt Civic Foundation since 2010.

What has mathematics meant to your career?

Mathematics has always been my passion. For me, it is an understanding of the world, of companies and of language: mathematics describes much more clearly than words. This is also true the other way around – what I have grasped and formulated mathematically can be put into words very succinctly. Mathematics naturally trains analytical skills and logical thinking, but also the ability to recognize and focus on crucial developments in complex systems. This has helped me at all my stations, regardless of the industry in question, even though I did not embark on a classical (engineering) scientific career.

What role does mathematics play in your institute/company? And more generally in the industry you work in?

Honestly, it has to be said that the mathematical questions at our company rarely go beyond a decent school level. We don't develop mathematical tools, and only in a few areas, such as power purchasing or grid control, are more complex simulations used.

Nevertheless, mathematics plays an important role for HEAG, both in our function as the parent company of the HEAG Group and in the investment management of the City of Darmstadt, the city of science. Reports on the economic situation, investments, planning figures and forecasts, personnel matters, purchasing of goods and services, donations, sponsoring, and the processes of digitization of the working world - we need mathematics to understand and implement all these examples. It is indispensable to be able to completely penetrate these topics and also explain them in simple terms - and this is where the precision and focus of mathematics helps. All this can be transferred from HEAG to the entire industry.

Fortunately, a recent development is that we increasingly have mathematicians in management positions. Where ten years ago it was more common to find lawyers and business economists, today mathematicians are taking on the responsibility.

What does a typical workday look like?

My typical workday is very diverse. Appointments with the owners of HEAG – these are mostly people with political responsibility – and the managements of our portfolio companies are just as much a part of it as work in supervisory committees and internal coordination rounds with departments or specialist areas. Then there's answering e-mails, reading reports, and so on. Due to the current situation, the vast majority of these discussions currently take place digitally - which would also be impossible without mathematics.

Are, from your point of view, mathematicians active enough to build bridges between university and economy/society?

On the one hand, I observe the positive development that more mathematicians apply for positions in business management and also find themselves in these functions.

On the other hand, mathematicians often still find it difficult to express themselves in such a way that not only specialist audiences are taken along. Together with the Darmstadt Civic Foundation, for example, we have been organizing a children's lecture at Darmstadt Technical University for children of elementary school age for many years. While experts from other natural sciences captivated the children with a fireworks display of impressive simulations and experiments, a really good professor of mathematics did the entire lecture with blackboard and chalk.

In addition, I think mathematics also has an image problem to some extent. I have experienced several times that mathematical congresses were opened by female politicians or politicians who then „pleased themselves" that they never understood math in school. Such situations are unworthy of politicians and should be called exactly that.

. Mathematicians could also take more of a stand on current issues of socio-political importance. For example, in the current COVID discussion, I miss eloquent mathematicians who comment on the situation and the resulting development variants.

What advice do you have for young people regarding mathematics?

Mathematics is the foundation for all engineering, natural science, social science and mathematics courses of study, as well as for numerous apprenticeship professions. But developing an understanding of numbers and mathematical relationships helps in almost all areas of life. In everyday life, of course, this begins with the rule of three, with quantities when cooking or with discounts in the supermarket. Mathematical understanding helps to better understand the processes of ubiquitous digitization. Just think of Google's search algorithms or QR codes. Encryption processes – whether for cell phones, computers or mobile payments – are also unthinkable without mathematics. It is good to be able to fundamentally understand and classify them. I can only recommend to anyone and everyone to get involved with mathematics and also to bite through it sometimes when it gets a bit more difficult. The better you know mathematics, the better you can explain the world to yourself.

Figure 2: Dr. Markus Hoschek

3 interview with Dr. Margrit Klitz

Dr. Margrit Klitz initially decided to study physics, but quickly took a liking to mathematics as a major. Today, she works at the German Aerospace Center in Cologne. She uses the language of mathematics in high-performance computing to investigate physical problems using high-performance computers.

A brief summary of your career

After high school, I didn't quite know whether to study medicine or physics. With medicine, I was attracted by the idea of being close to people, being able to help them, and being able to work anywhere in the world. A connection arose through medical physics, which fascinated me very much. I decided to study physics at the University of Bonn, but soon realized that I found mathematics more exciting. Therefore, I changed after one semester and studied mathematics with a minor in English studies. A rather unusual combination, but languages always came rather easily to me, and studying English probably couldn't hurt.

After my intermediate diploma in mathematics, I specialized in flow simulation, This was another big step in the direction of physics: How do drops behave on surfaces? How does the flow of water in a weir behave? How do you flow around textiles to produce fiber composites? And computer science: How can I calculate all of this in the computer and use it to create compelling images and scientific videos?

The field of flow simulation is large and much is still unexplored, so after my thesis I decided to do a PhD. This then took longer than expected because during this time my three daughters were born. It is not always easy to find a perfect time to plan a family, but I would do it all over again in retrospect.

After finishing my Ph.D., I stayed at the university for a few more months as a post-doc, before I took up my position as a research associate at the German Aerospace Center (DLR). As deputy head of department and group leader, I am now close to people again: I don't just work for myself, but also with the goal of making sure our employees feel valued and, in turn, can achieve their scientific goals with as little outside interference as possible.

I also returned to the University of Bonn: I held a practical programming course on „Numerical Flow Simulation“ for one semester and gave an introduction to the programming language „C“ for the fresh math students.

What has mathematics meant to your career?

As a mathematician, you learn to solve problems of all kinds analytically and with a lot of perseverance. This has given me a lot of confidence in my abilities. Since I didn't want to follow the classic path of a career in banking or insurance, but also didn't want to work purely in science, I quickly realized that I would have to learn a lot outside of university, but I'm not afraid of that. On the contrary, I believe that I will enjoy a job until there is no more opportunity to learn something new. For me, being a mathematician means biting down hard and not letting go until the problem is solved.

What role does mathematics play in your institute/company? And more generally in the industry, in which you work?

The „High-Performance Computing" department of the Institute of Software Technology, where I work, specializes in software problems, that can no longer be computed on one computer alone. Instead, many computers are used together, with a lot of power and communication between them, to solve these big problems. I was already familiar with this from flow simulation at university. The wide range of projects at DLR in the fields of aeronautics and space, but also energy, transport and security, make for varied scientific projects.

All of these projects are usually based on physical problems (Why does an airplane fly? How can I build faster trains? What happens in 3D printing?). These can be described by a physical model. The language in which this description is made is mathematics, and the solution to the problem on a computer is numerics or numerical simulation, which is also a branch of mathematics. So if you speak the language of mathematics, you are able to describe and, in the best case, solve such physical problems. The exciting thing about DLR in particular is that we have scientists from a wide variety of fields, such as engineering, psychology, physics and computer science. All of them bring expert knowledge to the table, which we can use in a targeted manner to improve the mathematical models and lead to a better solution.

Are, in your view, mathematicians active enough to build bridges between university and business/society?

Not every mathematician wants to do applied work. Some solve abstract problems, provide complex proofs, and deal with problems that are not application oriented. This helps us understand more and more connections in mathematics.

Many mathematicians work in insurance companies and banks, where they apply their knowledge in a targeted way, to strengthen and support the companies. In industry, numerical simulation replaces costly physical experiments.

We are addressing the major societal challenge of the current Corona pandemic using mathematical models. We want to predict how the pandemic will evolve for individual

Figure 3: Dr. Margrit Klitz

regions in Germany, how mobility between regions affects the pandemic, and which non-pharmaceutical measures have a large effect.

What advice do you have for young people regarding math?

If you enjoy puzzling, have stamina and enjoy mathematical problems, then you should study mathematics. Later on you can specialize further, e.g. you can orient yourself more in a physical or informatics direction. When choosing a career, you will be needed everywhere where you can support problems with your analytical approach.

4 Interview with Prof. Dr.-Ing. Stefan Kurz

„Tomorrow's products cannot be developed with yesterday's methods.“

A brief summary of your career

Prof. Dr.-Ing. Stefan Kurz (1966), studied electrical engineering at the University of Stuttgart, diploma 1992, doctorate 1998. Various professional positions at Robert Bosch GmbH and in the academic environment, among others Professor for Theoretical Electrical Engineering and Numerical Field Computation at the (then) University of the Federal Armed Forces Hamburg. „Finland Distinguished Professor“ at the (then) Tampere University of Technology (2010-2014).

Today working as Senior Chief Expert for „Hybrid Modeling“ (Physics and Data) At the „Bosch Center for Artificial Intelligence“ (BCAI).

. Simultaneously cooperating professor for mathematical modeling of electromagnetic fields at Darmstadt University of Technology.

What has mathematics meant to your career?

Mathematics has been a common thread throughout all stages of my career, which can be seen, for example, in the orientations of my professorships.

Even as a high school student, I was fascinated by mathematics, especially the realization that mathematics is a kind of „lingua franca“ of the physical world. For this reason, I chose the classical advanced courses of mathematics and physics in high school. I decided to study electrical engineering because I was curious about its applications. But already during my undergraduate studies I was tempted several times to switch to mathematics – until in my main studies the theory of electromagnetic fields and waves was added. This mathematically driven subject is with me to this day, for example for the design of superconducting magnets of the „Large Hadron Collider“ (LHC), the large particle accelerator of the European Organization for Nuclear Research CERN in Geneva.

In addition to the academic side, the application of mathematics also plays a crucial role in my professional environment. Today, I work at the „Bosch Center for Artificial Intelligence“ where, in addition to strategic issues, I deal with „hybrid modeling“. At its core, this is about coupling physics-based models with data-driven models, which is fundamental for Bosch. All the existing knowledge about the products and their cause-effect relationships from more than 130 years of company history comes predominantly from physics. This knowledge is integrated into the world of machine learning and artificial intelligence through „hybrid modeling“. Moreover, such „hybrid models“ help make artificial intelligence safe, robust, and explainable. This is a basic requirement for applying artificial intelligence in an industrial context.

The „hybrid models“ are equivalent to mathematical models, from different mathematical fields. For example, stochastic processes from the probability theory are coupled with differential equations.

What role does mathematics play in your institute/company? And more generally in the industry in which you work?

Mathematics plays a central role in product development at Bosch. As a leading international technology and service company, Bosch is represented in various industries. Its activities are divided into the four business sectors Mobility Solutions, Industrial Technology, Consumer Goods, and Energy and Building Technology. I would like to give two examples.

Permanent innovation in development methods is just as important as product innovation. A key element in today's product development is simulation. This requires physical and mathematical models from the real world. Equally necessary are efficient numerical solution methods that allow questions to be asked of the models on the computer, so to speak.

Figure 4: Prof. Dr.-Ing. Stefan Kurz

You may have heard of the so-called Moore's Law, after Gordon Moore, the co-founder of Intel. It states that the performance of computer hardware doubles empirically about every 18 months. What is less known is that a similar exponential development can be found in the field of mathematical algorithms, for example for the solution of the steady-state heat conduction equation. From the Gaussian elimination method, which you may know from school, to the modern multigrid method, a performance gain by a factor of hundreds of thousands has been achieved over almost 30 years[1]. Isn't that impressive? That's what fascinates me about applied mathematics.

Such performance improvements are essential in the age of digital transformation, where we see a rapid increase in networked and automated systems. In light of this increasing complexity, tomorrow's products cannot be developed using yesterday's methods.

My second example comes from the field of electric vehicle drive motor development. There, mathematics is the key to productivity and functional improvements in development. With so-called multi-objective optimization, design parameters are systematically varied so that tens of thousands of their variants can be processed on a high-performance computer. In this way, we design components at their physical limits, taking into account market requirements and resolving inconsistencies in design objectives.

Are, in your view, mathematicians active enough to build bridges between university and Business/Society?

In recent years, I see various initiatives to strengthen the important role of mathematics in schools and society, which I am very pleased about. With my cooperative professorship at the Technical University of Darmstadt, I myself am a bridge builder. Exciting application examples from industry and research make it possible to illustrate the theoretical principles in a comprehensible way. This creates a fruitful dialog for both sides. In joint discourse, new thought patterns emerge for application in research and industry. In addition, a valuable network has grown over time, which also offers interested students a great opportunity for professional orientation.

I also find the exchange between the generations to be a real enrichment. For me, a course is successful when I can form a kind of learning community with the students, from which all participants leave richer than when they enter. This enables me to keep my finger on the pulse of the times.

What advice do you have for young people regarding mathematics?

If you discover a mathematical interest in yourself, be curious and pursue it! My parents' generation still believed that if you were a person who successfully studied math, you either had to stay in college or go to an insurance company.

Far from it – with a good and solid math background, an incredibly wide variety of careers are open to you. One example is mathematician and former research minister Johanna Wanka.

Be bold. It's easy to say that it's „cool" to have been bad at math in school. Honestly, that's too short-sighted for me! Mathematics allows for a very precise and compact representation of facts. It was the mathematization of science – including logic – that

[1] OECD Global Science Forum, Report on Mathematics in Industry 2008, p. 8

made breakthroughs and the creation of clarity possible in the first place. Without mathematical knowledge, in my opinion, participation in culture-shaping developments is not possible.

To meet future societal challenges, industry needs outstanding talent from universities and colleges. This is especially true for applied mathematics, computer science, natural sciences and technology (STEM). This involves grasping the essence of a problem mathematically and applying suitable solution methods. We need so-called T-shaped competence profiles, i.e. a secure and deep mastery of one's own field, together with the ability to network with adjacent domains and with application areas. Be open-minded and look forward to getting to know the many faces of mathematics.

5 Interview with Prof. Dr. Volker Mehrmann

„Mathematics as the engine of business, business as the engine of mathematics“

A brief summary of your career

Volker Mehrmann completed his diploma in mathematics in 1979, his state examination in mathematics/physics in 1980, his dissertation in 1982, and his habilitation in 1987 at Bielefeld University. He spent research years at Kent State University 1979-1980, University of Wisconsin 1984-1985, and IBM Research Center Heidelberg 1988-1989. From 1990-1992 he held a professorship at RWTH Aachen University and he held the chair of Numerical Algebra at TU Chemnitz from 1993-2000. Since 2000, he has been a university professor of numerical mathematics at TU Berlin. From 2008 to 2016, he was the spokesperson of the DFG Research Center MATHEON and the Einstein Center ECMath. He was president of the Society for Applied Mathematics and Mechanics (GAMM) 2011-2013 and is president of the European Mathematical Society (EMS) since 2020.

His research interests include numerical mathematics, scientific Computing, applied and numerical linear algebra, control theory, and Dynamical (in particular Differential Algebraic) Systems, and in recent years Energy-Based Mathematical Modeling in particular using Port-Hamiltonian Systems.

What has mathematics meant to your career?

Mathematical structures and their use in the development, analyzability, and implementability of efficient numerical methods for industrial practice, has been and continues to be the driving force behind my research and teaching throughout my career. It has turned out over the years that this is not only a very successful strategy for constructing efficient and robust numerical methods, but also a wonderful motivation for basic mathematical research. As an example, the developments on structure-preserving linearizations of matrix polynomials, born from an industrial problem, this question has initiated a whole new theory with hundreds of mathematical publications. An extensive proof can be found in the books:

- M. Grötschel, K. Lucas und V. Mehrmann (Hrsg.), *Produktionsfaktor Mathematik. Wie Mathematik Technik und Wirtschaft bewegt*, acatech DISKUTIERT, acatech-Deutsche Akademie der Technikwissenschaften, Springer Verlag, Berlin, Heidelberg, 2008.
- P. Deuflhard, M. Grötschel, D. Hömberg, U. Horst, J. Kramer, V. Mehrmann, K. Polthier, F. Schmidt, C. Schütte, M. Skutella, and J. Sprekels (Editors), *Matheon, Ma-*

thematics for Key Technologies Series in Industrial and Applied Mathematics 1, EMS Publishing House, Zürich, Switzerland, 2014.

What role does mathematics play in your institute/company? And more generally in the industry, in which you work

Since I am a university teacher engaged in research and teaching, mathematics is a central part of my life.

Figure 5: Prof. Dr. Volker Mehrmann

What does a typical working day look like?

Preparing and giving lectures and seminars, supervising bachelor's, master's and doctoral theses, academic self-administration, bureaucracy with the university administration, applying for and carrying out research projects, discussing research discuss research questions, work actively on mathematical problems, prepare results for publication, write publications, prepare and give lectures, review publications, research projects and theses, give advice in supervisory committees of research institutes and much more.

Are, in your view, mathematicians active enough to build bridges between university and Business/Society?

There is considerable pent-up demand here. Many mathematicians:in do not want to or cannot tackle an application of their research results to business or society. This is definitely legitimate, because basic theoretical research is very important, because without it, mathematics cannot develop further and, in the long run, applications cannot be made. However, there is often a lack of communication about the results obtained to build bridges and often I experience a certain arrogance towards the colleagues who build these bridges. This is in no way justifiable.

What advice do you give to young people regarding mathematics?

If you like math and intellectual challenges, it's the greatest thing you can do. Mathematics permeates our society in all areas, mathematicians are needed everywhere, career opportunities are excellent. But it should give you pleasure, then it also makes you happy.

6 Interview with Dr. Michael Meister

A brief summary of your career

I have been very attached to mathematics from my school and college days into my professional and political life. At school I had great teachers who taught us students science with great enthusiasm. I then chose mathematics as my first examination subject in high school; so it was actually a matter of course to take mathematics with a minor in computer science during my studies at the then TH Darmstadt. My PhD thesis deals with the „investigation of the local truncation error for a linearized one-step method for the numerical solution of systems of quasilinear parabolic differential equations“.

After my military service, I then worked for five years as a flight dynamics expert at the European Space Agency Operations Center (ESA/ESOC) in Darmstadt. There I was responsible for orbit determination and attitude control of research satellites. Mission planning is used to define suitable time windows for the deployment of scientific instruments. In 1994 I was elected for the first time to the German Bundestag. There I was, among other things, Chairman of the Finance Working Group and later for a long time Deputy Chairman of the CDU/CSU parliamentary group for the areas of economy, small and medium-sized enterprises, finance and budget. From 2014 to 2018, I was Parliamentary State Secretary at the Federal Ministry of Finance, since 2018, I have held the same position at the Federal Ministry of Education and Research (BMBF).

What has mathematics meant to your career?

Above all, mathematics has taught me the importance of problem analysis and a structured approach to working out solutions. In mathematics, you have to make short, precise statements and always consider the entire solution space. And based on this analysis, you then have to choose the most appropriate solution path to achieve the goal. In other words, as a mathematician, you don't rule out any solution from the outset. I think that's why mathematicians communicate differently than non-mathematicians.

My knowledge of mathematics was essential to my job at ESOC. There I had to apply mathematics to concrete application fields in practice. And also in my life as a politician, mathematical working techniques and a logical approach to problems have helped me a lot. Especially in financial policy, which I dealt with for more than 15 years, my basic mathematical understanding was very helpful. This allowed me to easily work on budgetary or fiscal issues and understand complex processes such as sustainability analysis of public finances.

What role does mathematics play in your ministry? And more generally in the industry in which you work?

The BMBF attaches a central role to mathematics in science and research. Mathematics is needed to answer the most important social questions of our time - i.e., questions about climate protection, digitization, sustainability or even the origins of our universe. Mathematical solution concepts, models and methods make it possible to make quantitative predictions and find resilient answers. Mathematics is the basis for all natural sciences and technologies and thus also for the development of innovations. The BMBF contributes to the funding of numerous large-scale facilities for basic research in the natural sciences, which require state-of-the-art technologies and often reach an impressive complexity and size. Examples include the FAIR accelerator facility, the German Electron Synchrotron DESY or the European X-ray laser research facility XFEL.

Mathematics also plays an important role in education - the second policy area for which the BMBF is responsible - in the canon with the other natural sciences. In order to remain at the forefront of global economic competition, we need well-trained specialists. That's why the BMBF – but also me personally – attaches great importance to promoting the next generation of STEM students. In the MINT Action Plan, the BMBF has bundled its numerous funding programs to strengthen STEM education along the education chain. One central measure, for example, is the nationwide expansion of extracurricular STEM offerings for children and young people. In addition to the well-known school competitions such as the Math Olympiads or Jugend forscht, these STEM clusters help to get young people at a crucial age interested in mathematics, computer science and the natural sciences and to awaken their interest in a STEM profession.

Figure 6: Dr. Michael Meister

What does a typical working day look like?

What I like about my job is that every working day is different, depending on whether I am in Berlin during the weeks when the Bundestag is in session or whether I have appointments in my constituency of Bergstrasse in Hesse.

In Berlin, my work week is very well-paced: on Monday and Tuesday, various party committees meet; on Wednesday mornings, I attend the meetings of the Bundestag Committee on Education, Research and Technology Assessment; and the rest of the week, I represent the BMBF in the plenary sessions of the Bundestag. In addition, I work on parliamentary matters in my office, give speeches at conferences, or talk to members of parliament on the phone about education and research policy issues. A working week can have up to 70 working hours!

In my constituency, I hold discussions with citizens, representatives of local politics, companies or associations and attend to their concerns. And I enjoy spending time with my family!

But one thing is for sure: Thanks to my work as a member of parliament, I meet a lot of interesting people and am always dealing with new tasks and issues. And I can really make a difference, i.e. help citizens with their concerns and advance important social issues.

Are, from your point of view, mathematicians active enough to build bridges between university and economy/society?

In my view, interdisciplinary collaboration and the transfer from science to application are crucial. Only in this way can the potential of mathematics be fully exploited. The BMBF supports this, for example, with the funding priority „Mathematics for Innovations“. With around 5 million euros per year, application-related mathematical research is funded that contributes to quality of life and value creation in Germany. Projects are currently underway in the areas of health, energy transition and digitization. The range of topics is very broad. For example, mathematicians are working on approaches to improve medical care in rural areas, to make it easier to feed renewable energies into the power grid while maintaining grid stability, or to find solutions to the small-data problem in tumor diagnostics.

But it is also important to communicate with society. Mathematicians should be visible and make the contributions of mathematics to ünderstanding and shaping the worldünderstandable. Here, for example, Prof. Beutelsbacher comes to mind, who created the Mathematikum in Giessen, the world’s first mathematical hands-on museum. Luminaries such as Prof. Scholze, who won the most important award for mathematicians, the Fields Medal, in 2018 at the age of just 30, also contribute to bringing the fascination of mathematics to a broad public.

What advice do you have for young people regarding mathematics?

To go through the world with open eyes, because mathematics is present everywhere in our daily lives. When pulling into a parking space, comparing prices when grocery shopping, using a navigation device, or folding a paper airplane.

Curiosity is an important driving force for us humans and especially for children and young people. Curiosity means being open to new knowledge and new experiences. Meeting the phenomena of the real world and the different facets of an application problem with curiosity is the first step in the search for solutions.

Young people should take advantage of the many opportunities that mathematics offers. The abstract thinking you learn in mathematics is relevant in many professions. That's why mathematicians have excellent job prospects. This will continue to be the case in the years to come. For young people, mathematics offers a wide range of opportunities to tackle exciting issues and help solve pressing societal challenges.

What do you think is the best way for basic research ideas to make the leap into application?

It is important to bring the different players together. Mathematics is a very powerful tool – with enormous diversity of applications. Awareness of this strength is one of the prerequisites for seeking solutions in mathematics in the first place. On the other hand, application problems also challenge basic research and provide impetus for its further development.

In the BMBF funding priority „Mathematics for Innovation" mentioned above, researchers are working together with industry to develop new methods for application challenges. We also support this exchange between science and industry through numerous event formats. One aspect that should not be underestimated is the „transfer over heads". By trying their hand at application problems in the projects and contributing to new solutions, young scientists simultaneously strengthen their competence profile. It is no secret that graduates with practical experience are highly sought after on the job market - and with them, new mathematical ideas also enter companies.

Mathematics and its importance for our society have changed in recent years, especially due to digitalization and automation. What does this mean for young people?

Mathematics has existed for many thousands of years, but it is still evolving. Through digitalization and artificial intelligence, mathematics is penetrating more and more areas of work and life. The demand for mathematical methods and tools and their range of applications is enormous. It ranges from Internet search engines and social media to weather and climate forecasts and virtual product development to the effective control of energy grids, medical imaging procedures and individualized medicine. Mathematics makes it possible to evaluate complex data reliably and quickly and to gain new insights.

Digitization can help with this. In the coming years, numerous specialists and experts will be urgently needed in these areas to help shape this process of change.

7 Interview with Prof. Dr. Helga Rübsamen-Schaeff

A brief summary of your career

With a doctorate in chemistry, she first headed the Georg-Speyer-Haus Chemotherapeutic Research Institute in Frankfurt and built it up into a successful academic institution for basic medical research (1987-1993, focus: HIV, oncology). 1994-2006 initially head of virus research at Bayer AG, from 2001 head of all infection research (drugs against viruses and multi-resistant bacteria). 2006 founded the company AiCuris, management until 2015, continued work on drugs against infectious diseases. 2017 the approval of the first drug, Prevymis, - an extraordinary success for a young biopharma company. Prof. Rübsamen-Schaeff is currently a supervisory board member of AiCuris, a member of the supervisory board and the Board of Partners of Merck KGaA and E. Merck KG, respectively, and a member of the Supervisory Board of 4SC.

She was awarded the title of „Distinguished Woman in Chemistry and Chemical Engineering“ by IUPAC (International Union of Pure and Applied Chemistry), honorary membership of the Gesellschaft Deutscher Chemiker, she is a member of the National Academy of Sciences Leopoldina and in 2018 was awarded the German Future Prize, conferred by the President of the Federal Republic of Germany. In 2019, she received the Innovation Award of the State of North Rhine-Westphalia. In March 2020, she was awarded the Löffler-Frosch Medal of the Society of Virology.

What has mathematics meant to your career?

In drug development, you have to do a lot of modeling, for example, flooding and degradation of levels of drugs in the body, but you also have to use computer programs when you invent new drugs that calculate how exactly a drug fits into the structure you want to inhibit, or how to optimize it. And finally, one also needs a lot of statistics to verify research results, but also the results of clinical studies, to name just a few applications.

What role does mathematics play in your institute/company? And more generally in the industry you work in?

As stated earlier, mathematics plays a prominent role in all steps leading to a new drug.

What does a typical workday look like?

Now that I'm out of the lab, much of the work (in addition to traditional management tasks) is analyzing data, either those published by others or my own.

Are, from your point of view, mathematicians active enough to build bridges between university and business/society?

I am afraid that they are not, although mathematics is fundamental to so many areas of life, business, and research.

Figure 7: Prof. Dr. Helga Rübsamen-Schaeff

What advice do you give to young people regarding mathematics?

The same thing I advise for chemistry: „get involved in the study. Chemistry is such an important basic science, from which you can go into a wide variety of applications “. Likewise, mathematics is such an important foundation for describing all processes, that even after studying mathematics, a wide variety of careers and specializations are possible for your future professional life.

8 Interview with Prof. Dr. Anita Schöbel

At the Fraunhofer Institute for Industrial Mathematics, approx. 250 employees work on projects, that use mathematical models and methods to solve problems in business and society. Anita Schöbel is the director of the institute and a professor of applied mathematics at the TU Kaiserslautern. She talks about her enthusiasm for mathematics.

A brief summary of your career

After my mathematics studies at the Technical University of Kaiserslautern, I received my PhD in the field of discrete optimization. After my PhD, I worked for two years at the Institute of Industrial Mathematics, then returned to the university and habilitated there in mathematics in 2003. In 2004 I accepted a call to the Georg-August University in Göttingen, initially as W2 professor, after two external appointments in 2007 as W3 professor. I was Dean of Studies there, helped to establish the Simulation Science Center Clausthal-Göttingen and headed it for several years. I spent a research semester in Auckland (New Zealand) with my husband and our two children.

In 2019, I accepted an appointment at the TU Kaiserslautern and as head of the Fraunhofer Institute for Industrial Mathematics. For six years, I served on the board of the Society for Operations Research, in 2019/2020 as president. In Kaiserslautern, I am a member of the University Council, and at Fraunhofer, among other things, I am the spokesperson of the Strategic Research Field „Next Generation Computing“ and since 2020 in the scientific senate of the National Research Data Infrastructure (NFDI).

What has mathematics meant to your career?

Everything. Mathematics has been fun for me since I was in school. My enthusiasm arose when my teacher had me use quadratic addition on the blackboard to derive the formula for solving a quadratic equation. Nevertheless, the decision to study mathematics (I was also good in other subjects) was not easy for me. But I never regretted it. At the university, I continued to enjoy mathematics, that's why I stayed for my doctorate. And for the habilitation. I became a professor and eventually director of the Fraunhofer Institute for Industrial Mathematics ITWM.

The great thing about mathematics is two things: *first:* The logic and clarity of mathematics. What is proven is true. Mathematics brings order through beautiful structures, helps to see the essentials. *Second:* Mathematics is useful. Mathematics helps to treat cancer patients better, to save energy, to make traffic more efficient, to optimize production processes and materials. There are so many examples of how math improves our lives.

What role does mathematics play in your institute/company? And more generally in the industry, in which you work?

The Fraunhofer Institute for Industrial Mathematics (ITWM) in Kaiserslautern is one of the largest mathematical research institutes worldwide. With our nearly 250 employees and more than 60 PhD students, almost everything revolves around mathematics. We mainly deal with modeling, simulation and optimization. To this end, we continue to develop mathematics as a key technology, provide innovative impetus, support industry partners in the modeling and digitization of processes and products as well as the application and improvement of high-performance computing technologies, and we provide customized software solutions. Our customers are mainly found in the automotive, mechanical engineering, chemical, pharmaceutical textiles, medical technology, the computer industry, and the financial and insurance sectors. We were also able to provide mathematical support in combating the Corona crisis: For the prediction of the pandemic, in logistics support around vaccination, in the analysis of the soaking of masks or the dispersion of aerosols. The production of protective clothing can also be optimized using mathematical methods.

Figure 8: Prof. Dr. Anita Schöbel

Are, in your view, mathematicians active enough to build bridges between university and Business/Society?

Since its inception over 25 years ago, our Institute has had a bridge as an image to illustrate this very vision: We build bridges between the real world and mathematics. A real world problem is brought into the virtual world of mathematical methods by a mathematical model and solved there. The solution is then interpreted in the real world and successfully applied there (if everything worked out). This is exactly what we see as our task: To transfer findings from the university, new research results and methods into the real world, i.e. to apply them in business and society and to improve things there. Our motto is „Mathematics for a good future“.

But not only with us, also in many other areas I see activities of mathematicians to build such bridges. There are two applied mathematical research societies in Germany, the GAMM (Society for Applied Mathematics and Mechanics) and the GOR (Society for Operation Research). Both meet together with other disciplines, discuss practical projects and issues, and thus build bridges to the real world. In the GOR, for example, there are very active working groups, in which researchers regularly meet and exchange ideas with users from industry and business. This builds bridges on the research side. But bridge piers are also being built on the user side: Many companies have very well-positioned research departments staffed by mathematicians, who also travel to conferences and write research articles. I have the feeling that a lot is growing together here at the moment.

What advice do you have for young people regarding mathematics?

Here I have three pieces of advice at once:

Dare to study mathematics if you enjoy it! Mathematics is a cross-sectional science that plays a role in so many industries, that you will always find a good job, regardless of economic fluctuations.

Work with your peers and colleagues! Teamwork is important in mathematics, especially in applied mathematics. Together you are more creative, you have more new ideas. This is important if you want to make progress in mathematics.

Think outside the box! Mathematics can be applied in many disciplines if you get involved. Collaborative interdisciplinary work sometimes doesn’t fit the normal mold, but it is exciting, and when you get something done that is actually used, it is a wonderful result.

9 Interview with Prof. Dr. Andreas Schuppert

A brief summary of your career

After studying physics and earning a doctorate in mathematics at the University of Stuttgart, I went to work in industry at Central Research the then Hoechst AG, as a scientist in the Scientific Computing group. After almost 10 years at Hoechst and a part-time study of economics, I moved as head of the Process Engineering Software Group to the Central Engineering Department of the Bayer AG in Leverkusen and after two years took over the management of the mathematics group there in 2000, which we then called „Computational Solutions". The environment there was unique at the time: modeling and simulation had been introduced and had proven their value to the company, the current field of work we were able to tap into was hybrid modeling as a combination of machine learning and mechanistic, equation-based models. Here we were still developing current mathematical solutions and software and were able to apply them successfully in important applications, especially in materials research and development. Almost in parallel, we took up the topic of data mining at that time and also developed it into a mature field of application, especially for applications in troubleshooting in complex chemical and biotechnological production processes. Through these activities, especially the topic of data mining, I came into contact with the life sciences in pharmaceutical research and development as well as in crop protection research. Soon after, I was asked to take care of the then new field of systems biology and to build this up for Bayer, This opened up new opportunities for the development and adaptation of more complex mathematics for problems in pharmacology, clinical pharmacology, clinical development and diagnostics.

We pursued all these activities in cooperation with external partners, also from academia. This brought me into very close contact with applied mathematics, especially industrial mathematics, also on an international level. and was involved in academic committees. In 2007, I was appointed as a part-time professor for the subject „Data-Driven Modeling in Computational Engineering Sciences" at RWTH Aachen University, Faculty of Mechanical Engineering. The academic engagement was subsequently expanded piecemeal, so that in 2010 I exchanged my management position at Bayer for the position of Key Expert for Industrial Mathematics, the highest position in the then newly created expert career path.

Then in 2013, the „Joint Research Center for Computational Biomedicine" was established as the first private-public partnership in Germany between Bayer, RWTH Aachen University and the University Hospital Aachen under my leadership. In 2017, I moved to the Faculty of Medicine as Professor of Computational Biomedicine, where I have been working since 2019 after leaving Bayer AG to join 100%. We bridge the gap between computational sciences, especially mathematics, and the requirements of medicine.

Fields of work include. Hybrid Modeling, Digital Patient Models and also applications of quantum computing in biomedical problems.

Figure 9: Prof. Dr. Andreas Schuppert

What has mathematics meant to your career?

Mathematics has been extremely important to my career. Not so much the details, but rather the understanding of the mathematical methods underlying concepts and their strengths and weaknesses for particular applications. Only a deep understanding of these concepts allows you to, quickly develop problem-adequate workflows for solving complex application problems. Here, it is less important to work „methodsrein“ or to prove the existence of a solution, but on the clever combination of different methods, their implementation and reliability. Even if I can accurately estimate the error bounds, it does no good if they are far too large. This is not possible without a deep understanding of the respective mathematical concepts, which then also includes proofs. In this respect, we come full circle to „academic“ mathematics.

What role does mathematics play in your institute/company? And more generally in the industry, in which you work

Mathematics is a typical „enabling technology“ in the chemical and pharmaceutical industry and, as „mathematics“ such, does not play a major independent role. It gains its importance by enabling reliable solutions to very complex, often multidisciplinary problems with high value for the company. Ideally, it is not so much a matter of optimizing a little more of what has been established, but of opening up new options. This requires a very good understanding of the problem on the one hand and of the mathematical methods in question on the other. Often such projects between mathematics and application fail because of culturally conditioned problems of understanding, My studies in physics and later in business administration were very helpful in bridging this gap.

What does a typical working day look like?

It's very heterogeneous. At the beginning, I was involved in concrete research projects, with specific questions and tasks. These had to be structured, suitable workflows developed and implemented, and the application question itself solved and the solution validated. In addition to working with paper and computers, this requires a considerable amount of communication and an understanding of the working methods and problems of non-mathematicians, as well as reading the relevant papers. As a group leader, I then had to fulfill typical tasks of middle management, The mathematical part inevitably decreases and focuses on conceptual work. As a key expert, I then had to deal with more strategic issues, in particular the question of which mathematical topics could become important for the future challenges of the company, and how to deal with them. Here, the mathematical scope of activities has again expanded significantly. This also includes representing the company and the field of industrial mathematics in the academic community.

Are, in your view, mathematicians active enough to build bridges between university and Business/Society?

There has long been a great deal of activity on the part of mathematics from the university at the national and international levels, there is certainly no lack of it. The biggest hurdle, in my opinion, is cultural differences as well as a certain fear, of leaving one's comfort zone. This is true for all sciences, of course, but mathematicians are often rather introverted, which doesn't necessarily help. It takes a curiosity to get involved with new things, and also a certain tenacity and stubbornness, even when, even if the mathematical question for cracking the key point of the problem is not handed to you on a platter, but must first be painstakingly uncovered.

What advice do you have for young people regarding mathematics?

Do mathematics, do what fills them up in the process, and do it right. Most importantly, stay curious and open-minded. You will be amazed, how important mathematics is for mastering the complex challenges of our world and how much you can you can contribute as a mathematician.

10 Mathematics to Predict the Spread of a Contagious Disease**

The spread of the current COVID-19 pandemic can be described using a nearly 100-year-old model based on differential equations.

Dr. Maria Vittoria Barbarossa, group leader at the Frankfurt Institute for Advanced Studies, recalls the early days of the COVID-19 pandemic:

„As I write this post, exactly one year has passed since I began my work as a mathematician fighting the pandemic. SARS-CoV-2 had begun to spread rapidly through Europe, and in Germany there were about 400 confirmed infections. infections, plus a handful of people had died. I remember the phone call with my institute director, who said something like 'we are facing a European crisis, we have to get rid of this virus', You can model the spread of epidemics – do something!'. It was early March 2020 when we started with colleagues spread between Heidelberg, Munich, Jülich and Frankfurt – all in home offices, to collect the information and data on the new coronavirus, which at that time was still poorly organized. In a very short time, we set up mathematical models to represent scenarios for the spread of the virus based on the initial data and to make predictions for the impact of possible control measures (at that time, we did not yet have a vaccine available!). Over the course of the year, the models continued to be developed and refined to incorporate new insights and data, as well as to answer new questions - for example, about virus variants or the optimal distribution of vaccines."

Mathematical modeling with the SIR model

„Perhaps I'd better take a step back for the first time and explain in the first place what it means to set up a mathematical model for the spread of a contagious disease. Let us think of a disease, e.g. the classical flu, caused by a seasonal flu virus, which is transmitted from person to person. When we form a mathematical model, we should also consider what questions we want to answer through the model. For example, using data on the course of an influenza in previous years, we would like to predict how the influenza wave will be in the coming fall. At any point during the flu wave, for example, we want to know the number of people infected. We also want to know how many have recovered from the disease, because these people have developed full immunity to the disease after infection. For simplicity, we first do not distinguish between people who have recovered and people who die from the disease – even these people can no longer become infected with the disease and cannot infect another person. We then consider a geographic region and assume that movement into and out of that region can be neglected. In particular, we neglect births and deaths for the time we want to observe. Then we should define the essential model components (states and processes) that we absolutely need. The (constant) total population N in the region under consideration is then divided into three different groups:

- The susceptible individuals (S), i.e. those who have no protection against the disease and can be infected,
- The infected individuals (I) who are sick and contagious, and
- The recovered individuals (R) who are recovered (or deceased) and can neither infect nor be infected.

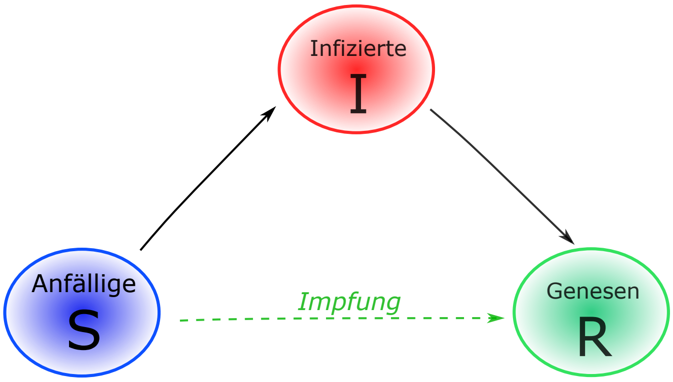


Figure 10: Schematic representation of the SIR model with transitions between the susceptible (S), infected (I), and recovered (R) groups of the population.

The change in the susceptible population at each time t is given by the derivative $S'(t)$ Suppose each infected individual has a fixed number κ of contacts per day, and that the probability of transmitting infection at any one contact is ρ. If we assume homogeneous mixing of the population, then for each susceptible individual the probability of becoming infected on any given day is given by $\beta I(t)/N$, where $\beta = \kappa\rho$. The newly infected individuals may no longer be in the susceptible group S and will instead contribute to increasing the infected population I. Thus we have that

$$S'(t) = -\frac{\beta S(t)I(t)}{N}.$$

Infected individuals $I(t)$ recover at a rate γ: that is, the *average duration of infection* is $1/\gamma$ days, i.e.

$$I'(t) = \frac{\beta S(t)I(t)}{N} - \gamma I(t).$$

The rate of change of the recovered group $R(t)$ is correspondingly

$$R'(t) = \gamma I(t).$$

These three differential equations (i.e. equations describing the rates of change of the observed states S, I and R) form the well-known *SIR model*, which was first proposed almost 100 years ago. With the help of numerical methods one can simulate the model on the computer and thus, for example, map the number of infected persons in time. In doing so, we can also see the typical course of a seasonal flu wave, which starts with only a few infected persons. The number of infected persons (and shortly after that the number of recovered persons) increases in time, while the number of susceptible persons decreases, until a peak in the number of infections is reached, and the epidemic slowly peters out (see Figure 11)."

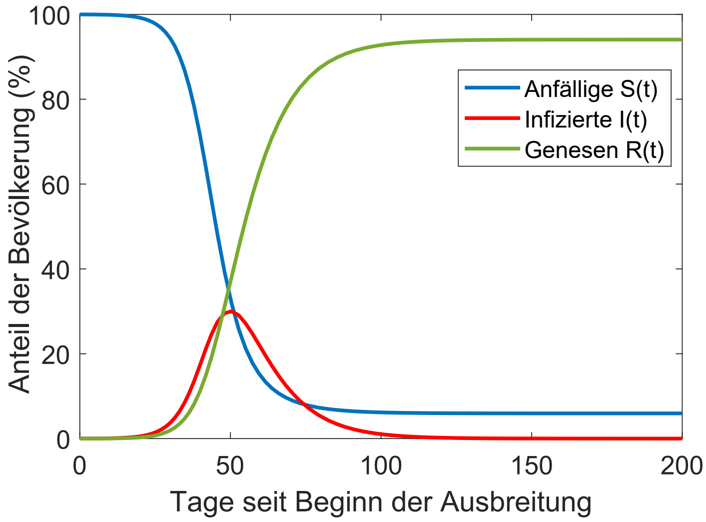


Figure 11: Typical course of an infection wave according to the SIR model.

The Basic Reproduction Number

„Furthermore, from this model, the *basic reproduction number*

$$R_0 = \frac{\beta}{\gamma}$$

determine. This number tells us how many people on average an infected person would infect, if that person were to arrive in a fully susceptible population. The basic reproduction number varies from disease to disease (even from model to model), for example, is between 2 and 3 for seasonal influenza and can be as high as 15-18 for measles. This number is particularly important because $1 - 1/R_0$ can be used to estimate the proportion of the population which is necessary for the so-called *herd immunity*. If we map the group of infected individuals in time for different values of R_0, we see that a reduction in R_0 leads to curves that have a low and shifted peak (see Figure 12)".

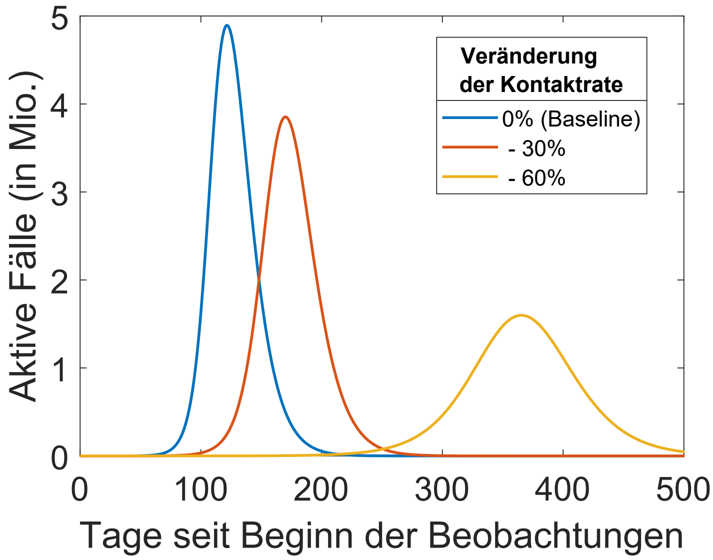


Figure 12: Effect of reduction of contact rate (or equivalently reduction of R_0) on the number of active cases. A low contact rate leads to a low and shifted peak.

Extensions to the SIR-Model

Returning to COVID-19 modeling, Ms. Barbarossa explains: „In essence, you can also use a relatively simple model like the SIR to describe the course of the pandemic. In practice, however, one would like to get more information from the model, so that one rather works with an extended version of the SIR model, where one also has the possibility e.g. non-detected infections (because mild, asymptomatic, or simply never tested), hospitalized cases, individuals requiring ICU stays, and even those who die from the disease. Let's look again at Figure 12 and think back to the control measures taken, which repeatedly led to contact reductions during the COVID-19 pandemic, then we understand what they were trying to achieve: Keep the number of infections at low levels for as long as possible, so that as many sick people as possible could receive good medical care. One can also integrate the effect of a vaccine in such a model (cf. Figure 10). Mathematically, by vaccinating, one shifts people from the susceptible group S to a group of vaccinated individuals, which in the best case behaves as well as the recovered group R, i.e., does not contribute to the spread of the disease.“

Dr. Maria Vittoria Barbarossa recalls her time as a student: „When I sat in the lecture halls of the Department of Mathematics as a freshman in the fall of 2003, I had no idea what I would learn in college and, more importantly, what I would do after I graduated with my bachelor's degree. In the 5th semester I was allowed to study as an Erasmus student at the TU Munich, where I heard for the first time about applications of mathematics in biology and life sciences. This fascinated me, I didn't want to leave the subject.

I quickly came to understand that the applications could go in very different directions. To give you an idea, in the years since my master's thesis, I've been privileged to work with microbiologists, biochemists, immunologists, epidemiologists, and people from intensive medical care. The challenge here is not so much the math. I find it much more exciting to get involved in other subjects and learn to define a common language in the process, to exchange ideas with colleagues from other disciplines."

11 The Reliability of COVID-19 Rapid Tests**

A mathematical theorem from the 18th century helps us to to investigate the significance of COVID-19 rapid tests.

COVID-19 rapid tests, also called „lateral flow tests“ (LFTs), are used to test employees before they start work, to allow visits to schools, stores schools, stores, restaurants, theaters, or to access 'body-related services' such as hairdressing.

But how reliable is such a test? How much can you rely on it? Let's take an example from the Robert Koch Institute[2]. In a low-incidence situation (5 out of 10,000 tested are actually infected), all 10,000 people take a test and 200 turn out positive, even though the person is not infected. This is then referred to as a 'false positive' rate of 200/10,000, or 2 %. So if the rapid test, only gives a false positive twice in 100 cases, what is the chance that one is actually infected if the test is positive? Is 98 % correct? The question was a bit mean because the person tested is missing some background information, without which, as we will see, this question cannot be answered.

A mathematical theorem from the 18th century helps us to answer this question. Bayes' theorem (also: Bayes' formula), named after the English mathematician Thomas Bayes (1701-1761), is used to calculate *conditional probabilities*. The formula of Bayes is in mathematical notation

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

Here, $P(A|B)$ is the (conditional) probability of the event A under the condition that B has occurred.

In our case (see RKI source, left side), 10,000 people are tested, selected at random. There are 5 out of 10,000 people infected with COVID-19. This is called the „preprobability“: the prevalence or background rate in the population, it is 0.05 % here. The rapid test correctly identifies 4 of them. There are 9,995 uninfected people and the rapid test correctly identifies 9,795 of them.

Properties of the rapid test: If one is infected, it will say so correctly in 4/5=80 % of the cases; if one is not infected, it will say so correctly in 9,795/9,995, or about 98 % of the cases. But this also means that although the rapid test will give the correct answer 80 % of the time, 200 people have been told that they are infected, even though they are not infected.

Back to mathematics and Bayes' theorem. As events, we define A: infected and B: positive test and obtain the probabilities

$$P(A) = 5/10,000, \quad P(B) = 204/10,000, \quad P(B|A) = 4/5.$$

[2] Robert Koch-Institute, Infographic: Understanding Corona Rapid Test Results, Feb. 25, 2021. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Infografik_Antigentest_PDF.pdf, pages 47–49

and we get

$$P(A|B) = \frac{\frac{4}{5} \cdot \frac{5}{10.000}}{\frac{204}{10.000}} = \frac{4 \cdot 5}{5 \cdot 204} = \frac{2}{102} \approx 2\%.$$

So, if you get a positive result, in this case the chance of actually being infected is only about. 2 %. Without knowing the prior probability, one does not know how likely it is that a result is false or true. As an exercise, one could use the result from the RKI source, right hand side; with this high incidence situation, every 10th person is infected and it gives quite different numbers!

Of course, these numbers are only true if the population is truly randomly tested. Do people use the rapid tests because they have have a suspicion (e.g., because they have symptoms), then the prior probability would be correspondingly higher.

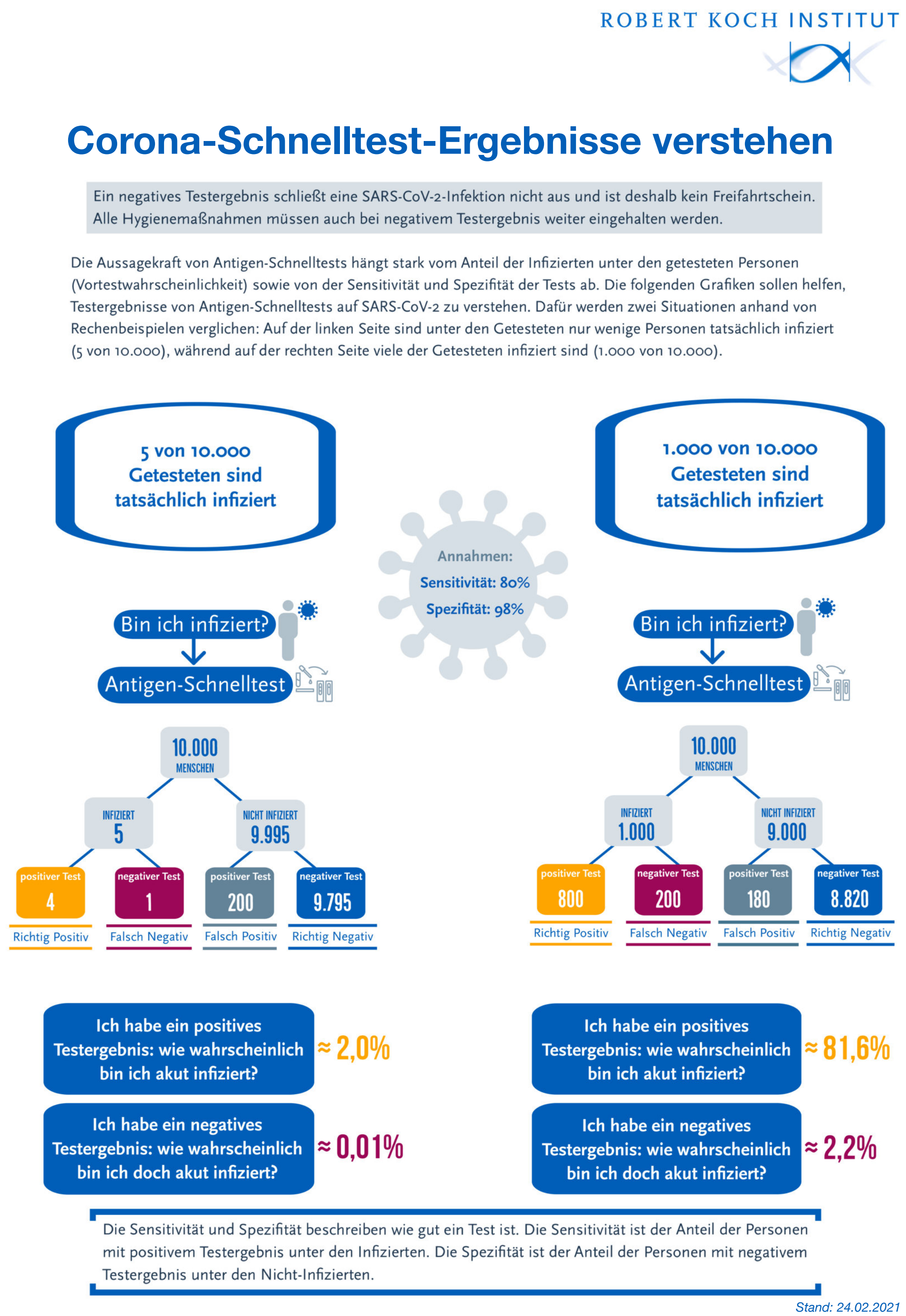


Figure 13: Infographic Corona rapid test, page 1 (Source: Robert Koch-Institute

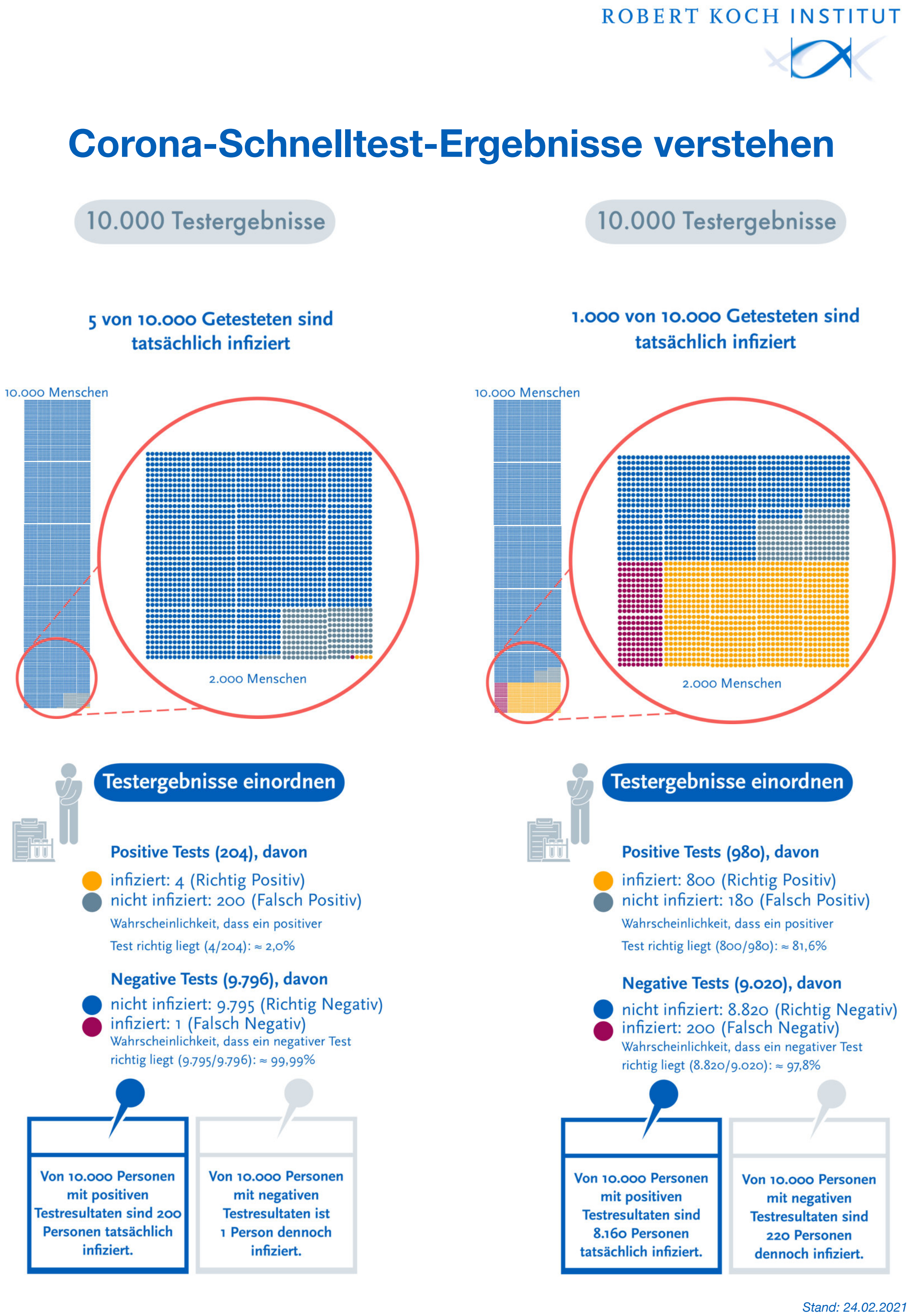
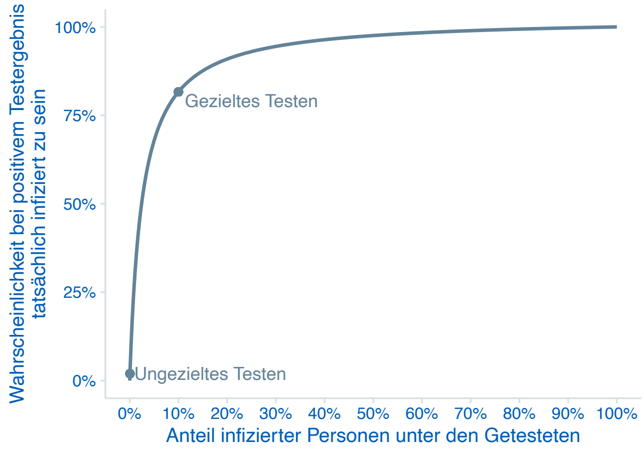


Figure 14: Infographic Corona rapid test, page 2 (Source: Robert Koch-Institute

ROBERT KOCH INSTITUT

Corona-Schnelltest-Ergebnisse verstehen



8,000 von 10,000 infizierten Personen werden korrekt als infiziert erkannt (Sensitivität = 0.8) &
9,800 von 10,000 nicht infizierten Personen werden korrekt als nicht infiziert erkannt (Spezifizität = 0.98)

Effektive Testansätze stehen im Zentrum der Bekämpfung von SARS-CoV-2. Für den Virusnachweis werden eine Vielzahl von Antigen-Schnelltests angeboten. Diese basieren auf dem Nachweis von viralem Protein in Abstrichen aus den Atemwegen. Antigen-Schnelltests können schneller durchgeführt werden als PCR-Tests. Im Vergleich zur PCR erkennen diese Schnelltests jedoch sowohl infizierte Personen schlechter (niedrigere Sensitivität) als auch nicht-infizierte Personen schlechter (niedrigere Spezifität). Die Aussagekraft von Antigen-Schnelltests hängt stark vom Anteil der Infizierten unter den getesteten Personen (Vortestwahrscheinlichkeit) sowie von der Sensitivität und Spezifität der Tests ab. Die Rechenbeispiele oben illustrieren den Zusammenhang zwischen dem Anteil der Infizierten unter den Getesteten, der Sensitivität und Spezifität der Tests und den resultierenden positiven und negativen Vorhersagewerten. Die angenommenen Werte für die Sensitivität und Spezifität der Tests sind großzügig angelegt. Der positive Vorhersagewert beziffert die Wahrscheinlichkeit, dass eine Person infiziert ist, wenn sie positiv getestet wurde. Der negative Vorhersagewert beziffert die Wahrscheinlichkeit, dass eine Person nicht infiziert ist, wenn sie negativ getestet wurde.
Wenn unter den Getesteten nur wenige Personen tatsächlich infiziert sind, dann sind positive Testresultate unzuverlässig. Wenn unter den Getesteten allerdings sehr viele Personen infiziert sind, dann sind positive Testresultate zuverlässig, aber die negativen Testresultate dafür weniger. Die Aussagekraft der Tests hängt vom Testansatz und der Verbreitung des Virus ab.

Wissenschaftskommunikation
Robert Koch-Institut
wiko@rki.de

Referenzen

- McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. Psychological Bulletin, 143(12), 1273. https://doi.org/10.1037/bul0000126
- Seifried, J., Böttcher, S., Oh, D.Y., Michel, J., Nitsche, A., Jenny, M.A., Wieler, L.H., Antão, E.-M., Jung-Sendzik, T., Dürrwald, R., Diercke, M., Haas, W., Abu Sin, M., Eckmanns, T., Hamouda, O., & Mielke, M. (2021). Was ist bei Antigentests zur Eigenanwendung (Selbsttests) zum Nachweis von SARS-CoV-2 zu beachten? Epidemiologisches Bulletin, 2021;8:3-9. http://dx.doi.org/10.25646/8040

Stand: 24.02.2021

Figure 15: Infographic Corona rapid test, page 3 (Source: Robert Koch Institute

12 Robust Incidence Numbers**

Mathematics can be used to stabilize incidence, but care must be taken to ensure transparency and understandability in order to gain widespread acceptance by the public.

In controlling the (non-pharmaceutical) interventions to the COVID-19 pandemic, the default guidance is the *7-days incidence* of confirmed new infections (sum of the last 7 days, per 100,000 population, based on reporting data in the period from the reporting date)[3]

$$\text{7-days incidence} = \frac{\text{number of confirmed new infections in the last 7 days}}{\frac{\text{considered population}}{100{,}000}}$$

is used. Thus, all reported new infections of the respective past 7 days are added and the sum is divided by the considered population (e.g., city, county, ...) and multiplied by 100,000.

To use this incidence value to control the pandemic, one could for example, set it as a target to reduce this incidence by 20 % each week[4]. Rather, however, absolute thresholds (e.g. 7-day incidence = 50, 100, etc.) have been used to add or relax measures (see, e.g. the German federal „emergency brake“ regulations[5]).

While some see this as a suitable early warning system, others cite the lack of robustness of the incidence figures. For example, with increased testing, this incidence would rise; however, it would then generally fall more rapidly as more cases are followed up. Similarly, incidence as an index implicitly assumes an equal distribution of infections in the tested group, which may not always be the case, as more people with a suspicion test themselves. What can be done?

First of all, the calculation of the incidence itself can be or comparable indices by a *linear regression* instead of simply taking the sum of the last 7 days to compensate for the so-called 'weekly seasonality' (weekend effect). Besides the stabilization, another advantage is that you get the value for 'today' and not the sum (i.e. a kind of average) of the last 7 days. Moreover, one could also estimate the chosen index value for some days in advance.

Linear Regression

In the following, we will go through the procedure of linear regression on a concrete example. For the example we change from the incidence to the so-called *doubling time*,

[3] Robert Koch-Institute: COVID-19 Dashboard, https://corona.rki.de/

[4] https://www.wa.de/politik/corona-notbremse-inzidenz-michael-meyer-hermann-markus-lanz-zdf-hamburg-90483279.html

[5] https://www.bundesregierung.de/breg-de/suche/bundesweite-notbremse-1888982

which is to be estimated on the basis of the trend in the data. At the end of April 2020, the trend in the data was already clearly decreasing (effective reproduction number <1 since the end of March 2020) and the doubling time became significantly longer compared to the estimates in the first weeks of the pandemic (where we were in the exponential-growth trend).

In the course of the Corona pandemic, Johns Hopkins University (JHU) reported the following data for Germany during the period April 21-26, 2020:

Date	18.04	19.04	20.04	21.04	22.04	23.04	24.04
I	143.342	145.184	147.065	148.291	150.648	153.129	154.999

Where I denotes the total number of confirmed infections.

As a simple approximation in the initial period of a pandemic (or the beginning of a wave of infection), an exponential growth

$$I(t) = I(t_0)\, e^{b\,(t-t_0)}, \quad t > t_0$$

with parameter b be assumed. Now, the *doubling time* T_2 on 24 April 2020 is to be predicted in this model as a measure of incidence. It is

$$\frac{I(T_2)}{I(t_0)} = 2 \quad \text{i.e.} \quad e^{b\,(T_2-t_0)} = 2.$$

Thus we obtain the doubling time $T_2 = t_0 + (\ln 2)/b$.

We now consider the logarithmized values of the infection numbers

$$L(t) := \ln I(t) = \ln I_0 + b\,(t - t_0)$$

and calculate by (linear) regression on a reasonable time scale polynomials, e.g. $p_1(t) = a_0 + a_1(t - t_0)$. We determine b and thus T_2 on April 24 by deriving the approximations for $L(t)$ and compare the result with that of the JHU (53.0 days)! Numbering the days in ascending order by $t_0 = 0, \dots, t_6 = 6$, we obtain for p_1 the problem $||A_1x_1 - v||_2 \to \min$, where

$$A_1 = \begin{pmatrix} 1 & \dots & 1 \\ 0 & \dots & 6 \end{pmatrix}^\top, \quad x_1 = \big(a_0, a_1\big)^\top, \quad v = \big(\ln I(t_0), \dots, \ln I(t_6)\big)^\top$$

Furthermore,

$$A_1^\top A_1 = \begin{pmatrix} 7 & 21 \\ 21 & 91 \end{pmatrix}, \quad A_1^\top v \approx \big(83.3772, 250.4968\big)$$

and we obtain as solution of *normal equations* $x_1 \approx \big(11.8719, 0.0130\big)^\top$. Deriving yields $b \approx p_1'(t_6) = a_1 \approx 0.0130$, where $T_2 = (\ln 2)/b \approx 53.1438$, which is the value of JHU.

If, on the other hand, a quadratic polynomial is used $p_2(t) = \hat{a}_0 + \hat{a}_1(t-t_0) + \hat{a}_2(t-t_0)^2$, after deriving $b \approx p_2'(t_6) = 2\hat{a}_2(t_6 - t_0) + \hat{a}_1 = 12\hat{a}_2 + \hat{a}_1 \approx 0.0146$, where $T_2 = (\ln 2)/b \approx 47.48$. This (better?) prediction deviates more strongly from that of the JHU. Thus, the JHU will probably also have used a linear fit with polynomial p_1.

Robust indices with delay

Efforts are made to modify the incidence value or a corresponding index so that it is more informative for the course of the pandemic and thus for the measures to be taken. A classic COVID-19 model has already been presented in Chapter 10. Now, while there are robust, reliable values, such as.

- admissions to a hospital (so called hospitalization).
- number of COVID-19 patients in intensive care units (patients requiring intensive care)
- Number of available intensive care units (ICUs), including the associated specialist nurses
- Number of persons who have died (of/with COVID-19).

However, these values have a long lag (e.g. 14 days) to the possible measures (in mathematics: 'control variables') and are therefore of limited use as a sole index.

One solution would be an index as a (weighted) average of the various individual indices, comparable to a portfolio of stocks. Also on the stock market, one would like to reduce the risk of being wrong by this diversification. One disadvantage is that this approach (with the socio-economic consequences) is less transparent for the population and the exact choice of the portfolio is too complex. and the exact choice of the portfolio is to a certain extent arbitrary.

Another way to make the incidence more stable to the number of testings would be to estimate and account for the *dark rate of new infections*. To do this, one extends the SIR model from Chapter 10 with additional compartments to account for robust data (see above) and incorporates a dark rate factor ω. Now fit the extended model to the data (hospitalization rates, death rates) and in this way find the factor ω that best fits the data and use it as an estimate for the dark rate[6].

The EPG Index

Another alternative is the *EPG index* (Effective Potential Growth), which we briefly summarize here from an article[7]. For this purpose, we start with the *infected-fatality ratio* (IFR), which is the proportion of deaths among all infected persons:

$$\text{IFR} = \frac{\text{number of people who died from COVID-19 in a period}}{\text{total number of COVID-19 infections in the same period}}.$$

Thus, the IFR estimates the risk of death among all infected persons (diagnosed, asymptomatic, untested).

[6] A. Kergaßner, C. Burkhardt, D. Lippold, et al, *Memory-based meso-scale modeling of COVID-19*, Comput. Mech. 66 (2020), 1069-1079. https://doi.org/10.1007/s00466-020-01883-5.

[7] Martí Català, David Pino, Miquel Marchena, et al.,*Robust estimation of diagnostic rate and real incidence of COVID-19 for European policymakers*, PLOS, January 7, 2021 https://doi.org/10.1371/journal.pone.0243701

First, we take an estimate IFR = 1 % (in Europe, 0.3–3%) and, taking into account the reported (relatively certain) numbers of deaths, we estimate the period *(mean) time to death* TtD to TtD=18 days (people die between 15 and 22 days after the onset of the first symptoms)[8]. The estimated number of infected individuals with COVID-19 E_t at time t, is:

$$E_t = \frac{d_{t+\text{TtD}}}{\text{IFR}},$$

where $d_{t+\text{TtD}}$ is the number of reported deaths at time $t+$TtD. This allows us to estimate the number of cases before TtD=18 days and compare the value to the number of cases detected before 18 days, giving us a diagnostic percentage.

However, this is an unrealistic lower bound because no one performs a PCR test or sees a physician on the first day of symptoms. Similarly, external factors, such as the availability of PCR testing, saturation of the health care system, etc. may delay diagnosis. This *delay to diagnosis* (DD, 4-14 days) corresponds to the time between the onset of the first symptoms and reporting by the health care system. This time is highly country-specific.

Authors Català, Pino, Marchena, et al. propose the following. general procedure for the calculation of the *percentage of diagnosis* (%D):

1. Using the above formula, TtD and IFR can be used to estimate the number of cases can be estimated using TtD and IFR (alternatively using a fitted model, see previous section).
2. Calculate the temporal correlation (relationship) between reported cases C_t and reported deaths d_t. The maximum of the correlation determines the time between (confirmatory) diagnosis and death (DtD, 4-14 days). The delay in diagnosis DD is calculated as DD=TtD-DtD.
3. Evaluation of the percentage of diagnosis %D based on estimated (1st) and reported (2nd) cases.

Once the diagnosis rate %D is known, it is straightforward to determine a real incidence that no longer depends on (country-specific) time delays to diagnosis. The level of diagnosis and real incidence is useful for policy makers. Because the diagnosis rate %D depends strongly on the country-specific value for DD, we also define the *delay-to-detection diagnosis rate* (DD-DR) as the diagnosis rate %D calculated using a different time delay between symptom onset and detectability for each country.

Using DD-DR, one can calculate how many undetected individuals have been added to the number of infected $I_{t,14}$ in the last 14 days. Thus, one can better estimate the 14-day attack rate. The attack rate A is the proportion of a population that becomes ill

[8] F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, et al, *Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study*, Lancet 395 (2020), 1054-1062. https://doi.org/10.1016/S0140-6736(20)30566-3

with the infection in a certain period of time[9]. Here's what we want to do. The 14-day attack rate at time t is defined as

$$A_{t,14} = \frac{\text{exposed susceptibles in a group who became ill in the last 14 days}}{\text{exposed susceptibles in that group overall, in the last 14 days}}$$

Instead of the *(effective) reproduction number* R_t estimated from extended SIR models, we use as an empirical alternative ρ_t, the number of new cases detected today N_t divided by the number of new cases detected five days ago: N_t/N_{t-5}. However, the wide variation in these quantities requires that we use averaged values:

$$\rho_t = \sum_{i=-n_d}^{n_d} \frac{N_{t+i}}{N_{t+\tau+i}}.$$

Specifically, because of the pronounced week seasonality, we use values over seven days ($n_d = 3$) and delay $\tau = 5$ days, which is approximately the time for infected individuals to develop symptoms, if they develop any, i.e., we take

$$\rho_{t,7} = \frac{N_{t-3} + N_{t-2} + N_{t-1} + N_t + N_{t+1} + N_{t+2} + N_{t+3}}{N_{t-8} + N_{t-7} + N_{t-6} + N_{t-5} + N_{t-4} + N_{t-3} + N_{t-2}}.$$

This rate is less than 1 when the number of new cases decreases and greater than 1 when the number of cases increases. Averaging processes (or better equalization, see above) can further stabilize this value $\rho_{t,7}$ and is denoted by $\bar{\rho}_{t,7}$.

Authors Català, Pino, Marchena, et al. propose the following daily EPG index:

$$\text{EPG}_t = \bar{\rho}_{t,7}\, A_{t,14}.$$

The EPG index is simply the multiplication of the (averaged) disease growth rate $\bar{\rho}_{t,7}$ by the estimate of the attack rate $A_{t,14}$, both evaluated at the correct time t in the recent past.

The worst case is one where both $A_{t,14}$ and $\bar{\rho}_{t,7}$ are large. This means that you have a large population with the disease and lots of spread earlier. The best situation is a low value of the velocity and a small number of active cases. A large number of $A_{t,14}$, with low $\bar{\rho}_{t,7}$ or a large $\bar{\rho}_{t,7}$ with low $A_{t,14}$ are potentially dangerous situations.

[9]The ECDC (European Centre for Disease Prevention and Control) uses the period of the last 14 days in this context, *Rapid Risk Assessment: Coronavirus disease 2019 (COVID-19) in the EU/EEA and the UK - tenth update*. 11th June, 2020. Available: https://www.ecdc.europa.eu/en/publications-data/rapid-risk-assessment-coronavirus-disease-2019-covid-19-pandemic-tenth-update.

13 Mathematical Modeling and Optimal Control of Epidemics***

Mathematical models are essential for predicting the course of epidemics and the effect of intervention measures. Optimal control theory allows efficient intervention strategies to be calculated directly.

Dr. Markus Kantner and Dr. Thomas Koprucki of the Weierstrass Institute for Applied Analysis and Stochastics, Berlin describe the application of optimal control theory to the COVID-19 pandemic:

The World Health Organization estimates that there are about 200 zoonoses, i.e. infectious diseases that jump from the animal kingdom to humans through adaptation of pathogens. Some of these new pathogens are harmless, while others can cause catastrophic epidemics. An important parameter to determine the transmissibility of a new infectious disease is the *basic reproduction number* $\mathcal{R}_0$. It indicates how many people an infected person will infect on average if no immunity at all prevails in the overall population. immunity prevails. The numerical value of the basic reproduction number is not a natural constant, but depends on both the biomedical properties of the pathogen and the degree of interconnectedness of society and its basic health status.

If the basic reproduction number is less than 1, the number of infected people in the population decreases over time and the infection event comes to a halt. If it is is greater than 1, there is an exponential increase in the number of infections („snowball effect"), see Figure 16. In the case of the COVID-19 pandemic, for example, estimates of the baseline reproduction number range from about 2.4 to 3.5, resulting in a virtually worldwide rapid exponential growth in the number of infections.

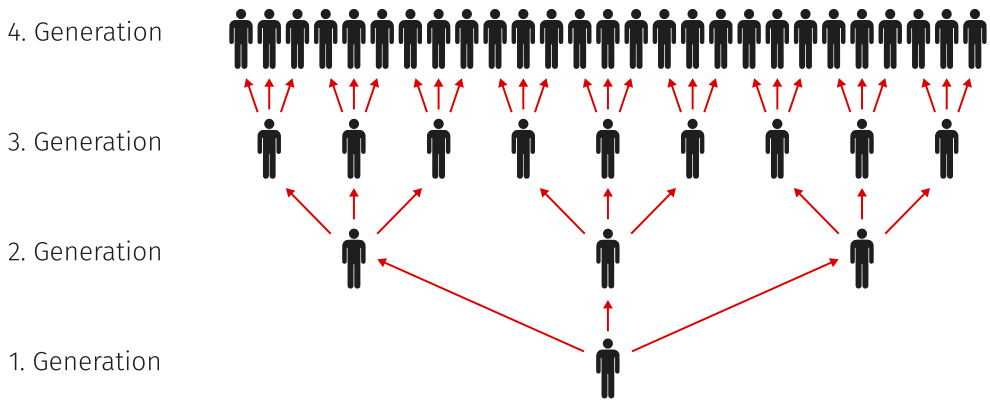


Figure 16: Exponential growth of an uncontrolled epidemic with basic reproduction number $\mathcal{R}_0 = 3$.

Nonpharmaceutical Interventions

For emerging infectious diseases, neither effective medical therapies nor vaccines generally exist at the beginning of an epidemic, neither effective medical therapies nor vaccines exist. To mitigate the exponential growth, non-pharmaceutical countermeasures must be taken in such a situation. The scope of such non-pharmaceutical measures is vast, ranging from intensified hand hygiene and the wearing of protective medical masks to „lockdown measures" such as imposing curfews, school closures, and banning large events. In addition, testing and quarantine of infected persons and their contacts can further reduce the *effective reproductive number* until it eventually falls below 1 and chains of infection break.

However, an exponential increase in the incidence of infection occurs again at any time after these intervention measures are removed, as long as no „herd immunity" prevails in the population as a whole (see the COVID-19 modeling chapter, page 40). If the pathogen cannot be eradicated, nonpharmaceutical measures thus buy time to develop drugs and vaccines in particular.

On the other hand, these measures may be associated with high burdens and costs for society. This concerns, for example, the economic damage caused by the closure of stores as well as the psychological effects of prolonged contact restrictions. Nonpharmaceutical measures must therefore be used in a coordinated to achieve effective containment of infection dynamics while minimizing socioeconomic damage. But how should such measures be be meaningfully coordinated and their impact predicted?

Mathematical Modeling: Understanding the Epidemic with Mathematics

Mathematical models can provide one answer to this question. Infectious disease dynamics can be modeled in a variety of ways. They range in scale from simple models (*differential equations*), which reflect the course of an epidemic for larger populations on average, up to very complex and detailed models, that describe infection events at the level of individual agents (*agent-based, stochastic models*).

The latter require a large number of parameters and data (e.g. contact networks, age structure, mobility data, health data, etc.) to make accurate predictions, whereas simple models can adequately reflect infection events with relatively few but well-fitted parameters. Mathematical models allow the effects of school closures and similar contact reduction interventions to be analyzed and evaluated in computer simulations.

We will illustrate the basic principle of non-pharmaceutical intervention measures using the example of the simple SIR model model, which we already learned about in the previous chapter (see the COVID-19 modeling chapter, page 40). In this model, the total population is divided into three groups: S (*susceptible*) for the initially healthy and infectious, I (*infectious*) for the infected and infectious, and R (*removed*) for the recovered and deceased. In the simulation, for a given time period, the contact rate β is reduced, although in this simple model it remains open by what specific measures this is achieved.

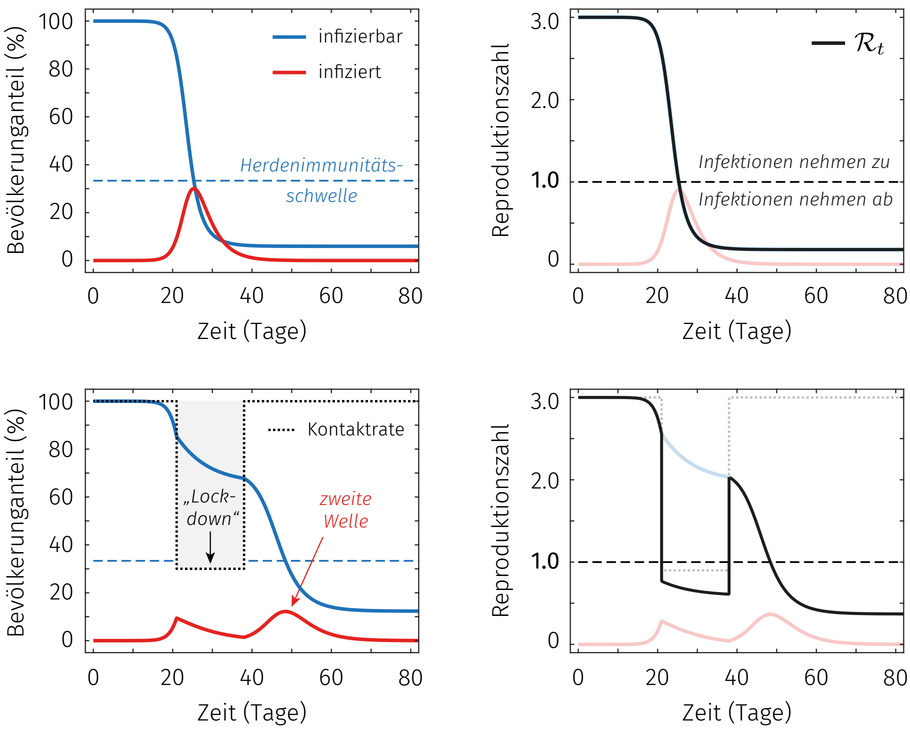


Figure 17: Top: Progression of an uncontrolled epidemic and associated effective reproductive number. After the proportion of infecibles has fallen below the herd immunity threshold (dashed line), infection numbers begin to decline. Bottom: simulation of a short lockdown by temporary contact reduction. After the measure ends, a second wave of infection occurs.

Figure 17 shows the effects of contact reduction interventions. Infections are shown to increase again after the measures are removed („second wave“), as long as the herd immunity threshold has not yet been reached (i.e., the number of infectious individuals is still too large, dashed line). Such „on-off interventions“ can also be switched arbitrarily often see Figure 18, whereby flaring waves of infection are repeatedly interrupted.

Theory of Optimal Control: Managing the Epidemic with Mathematics

The above example shows that in the case of simple on-off strategies, the epidemic keeps flaring up and further interventions become necessary. However, repeated shutdowns are most likely to lead to low acceptance in the population and the economy. But what could an intervention strategy look like that avoids these problems?

The *theory of optimal control* provides a systematic approach to calculating the ideal time course of the mean contact reduction, where both the socio-economic costs and other constraints can be taken into account. Such constraints include, the limited num-

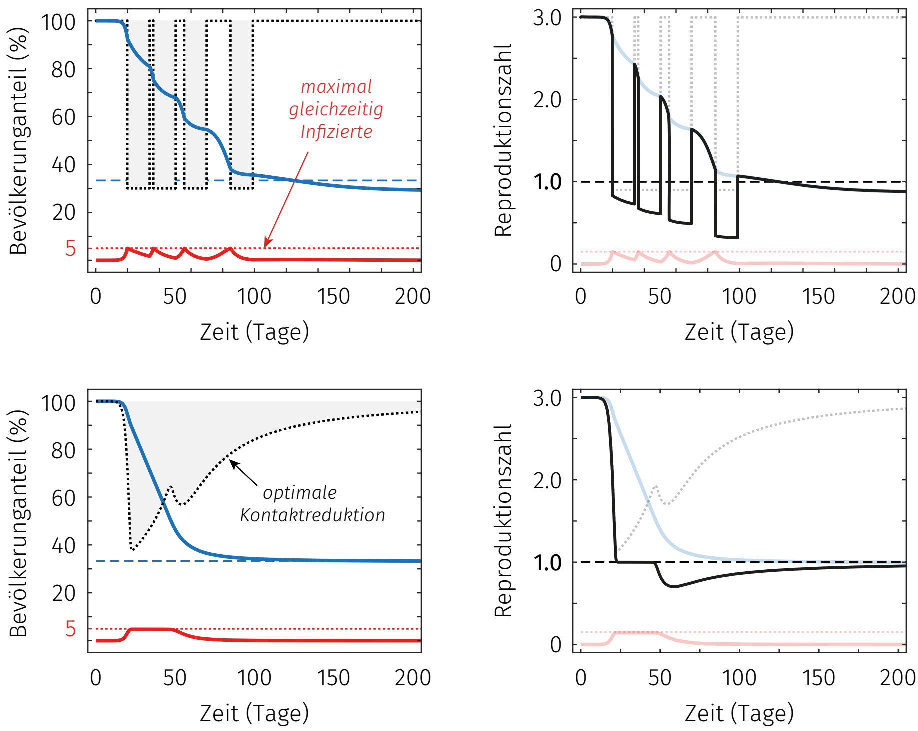


Figure 18: Top: On-off control of an epidemic with the constraint that the proportion of infected people in the total population should never exceed 5 %. Bottom: Optimally controlled contact reduction ensures that infection numbers never exceed the critical upper limit, while measures can be gradually relaxed.

ber of available intensive care treatment places, which must not be exceeded in any case. At the heart of the method is, in addition to the model for the spread of infection, a so-called *target functional*, which reflects both the goals of the intervention and its (abstract) costs. Methods from the *calculus of variation* can thus be used to derive conditions (in the form of mathematical equations), from which the optimal time course of the measures for contact reduction can then be calculated.

Figure 18 shows the course of an *optimally controlled* epidemic in the SIR model, in which the non-pharmaceutical measures are coordinated in such a way, that a certain upper limit of concurrently infected persons (at most 5 % of the population) is never exceeded. In this case, the measures can be gradually relaxed over time because the steadily increasing immunity in the population continuously allows more contacts without immediately reverting to exponential growth. The intervention ends when the herd immunity threshold is reached, so that subsequent waves of infection are excluded. The total socioeconomic cost of the optimally controlled intervention is significantly lower than in the case of repeated lockdowns (on-off strategy).

Similarly, such mathematical techniques can be applied to much more complex epidemiological models that can serve as a basis for policy decision making. Furthermore, vaccination campaigns and their optimal coordination can be planned using mathematical methods (see subsequent chapter on vaccination strategies, page 60). Mathematical modeling of epidemics is thus essential to obtain forecasts of future infection dynamics and to be able to initiate appropriate countermeasures at an early stage.

Incidentally, the same mathematical techniques of optimal control are also used in Chapter 31 to determine an optimal CO_2 tax.

14 Mathematical Modeling of Vaccination Programs**

Vaccination is the best way to prevent viral epidemics. But the wrong strategy in vaccinating the population can have disastrous consequences.

Prof. Dr. med. Helmut Brunner from the University of Düsseldorf and the Chair of Health Economics at the University of Wuppertal and Prof. Dr. Matthias Ehrhardt from the University of Wuppertal describe their experiences with mathematical modeling of vaccination in this chapter.

Vaccination programs against smallpox, diphtheria, polio, meningitis, measles, mumps, rubella, pertussis, and tetanus have proven to be very effective. Despite this, eradication of the target disease has only been successful for smallpox. Adult vaccination often proves less successful because switching to other serotypes of the pathogen (HPV vaccination) or variants (mutants) of the pathogen (antigenic drift or shift in influenza) can weaken vaccine protection.

Immunological Basics

Innate (nonspecific) and acquired (specific) immunity are available to defend against infectious diseases. The specific immune system with its capacity for „immunological memory" forms the prerequisite for immunization by vaccination. The main players of specific immunity are B lymphocytes, which can differentiate into plasma cells and produce antibodies, and T lymphocytes, which represent the cellular defense. The first contact of B cells with antigens of infectious agents produces predominantly IgM class antibodies, and the second contact produces IgG class antibodies. These are often detectable for life and ensure lasting protection via B memory cells. Since many antigens cannot activate B cells directly, cooperation of macrophages and B lymphocytes with T lymphocytes (CD4, CD8 cells) is necessary.

Active Immunization

In active immunization, specific antigens of infectious agents are administered intramuscularly or subcutaneously to induce antibody production. Usually, multiple vaccinations are given (boosters) to raise antibody titers and provide long-lasting protection. Live vaccines (attenuated infectious agents still capable of reproducing) and inactivated vaccines (component vaccines and whole germ vaccines) are used. New innovative vaccines based on genetic engineering (mRNA) have only been used for a few months and need to be tested for their practicality in fighting the SARS Cov-2 pandemic under routine conditions.

Vaccination programs aim to prevent illness and death from infectious diseases in children and adolescents. The substantial national and international success of prevention through active immunization has been widely documented. Figure 19 shows the example of a bacterial disease, purulent meningitis (meningitis) caused by *Haemophilus influenzae* Type b, and a viral disease poliomyelitis (polio), the decrease in disease incidence after the introduction of vaccination programs[10].

On the other hand, vaccination skeptics have claimed that the overall mortality rate, mortality, could be reduced only slightly by vaccination programs, since other measures, such as better treatment methods, anti-infectives, hygiene, nutrition, and other socioeconomic factors already explain the decline in mortality from infectious diseases. Therefore, mortality from the above diseases had already declined before the introduction of vaccination. The evidence that *Active vaccination* is one of the most important and effective measures of prevention from deaths in children and adolescents has been proven only a few years ago, among others, by studies from the Netherlands[11]. Mathematical models were used, among other things, to determine the years of life lost. The mortality of children and young adults up to the end of the 20th year of life decreased exponentially in the period from 1903 to 1992 (with a half-life of 19 years) and also in the period before the introduction of the first vaccination. However, the authors concluded that vaccinations against childhood diseases saved the lives of 6,000 to 12,000 children in the Netherlands in the 20th century. For example, regression models and time series of the following form were used:[12]

Let Y_t be the observed number of reported cases in month t, $t = \{1, \ldots, n\}$, and n be the total number of months before vaccination. Y_t follow a Poisson process.

$$Y_t \sim \text{Poisson}(\mu_t).$$

[10] U. Heininger, *Risiken von Infektionskrankheiten und der Nutzen von Impfungen*, Bundesgesundheitsbl. Gesundheitsforsch. – Gesundheitsschutz 47 (2004), 1129-1135.

[11] M. van Wijhe, S.A. McDonald, H.E. de Melker, et al., *Effect of vaccination programmes on mortality burden among children and young adults in the Netherlands during the 20th century: a historical analysis*, The Lancet Infectious Diseases 16 (2016), 592–598.

[12] M. van Wijhe, A.D. Tulen, H. Korthals Altes, et al., *Quantifying the impact of mass vaccination programmes on notified cases in the Netherlands*, Epidemiology and Infection 146 (2018), 716–722.

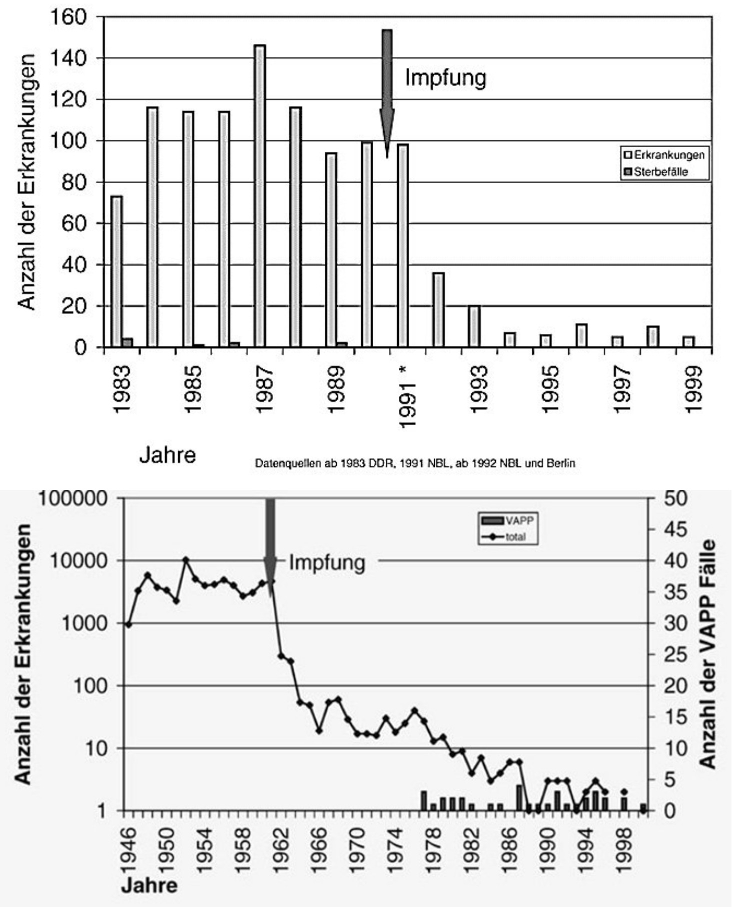


Figure 19: Top: *Haemophilus influanzae* Meningitis in the former GDR and the new federal states 1983-1999 from 1991 Federal Republic of Germany, bottom: Diseases of poliomyelitis, VAPP: Vaccine Associated Paralytic Poliomyelitis (Own representation according to Heininger 2004)

The regression model can be described as

$$\log(\mu_t) = \log(p_t) + \beta_0 + \beta_1 t + \sum_{j=1}^{k} \left[a_j \sin\Big(\frac{2\pi t}{12\tau_j}\Big) + \gamma_j \sin\Big(\frac{2\pi t}{12\tau_j}\Big) \right] + x_t,$$

where ρ_t is the general population of 0 to 20 year olds at time t given as „offset“ (the population at highest risk of infection), β_0 is the intercept, β_1 is the coefficient of the secular trend (transformed to indicate a percentage change: $\tilde{\beta}_1 = 1 - \exp(\beta_1) \times 100$). Seasonality and multi-year cycles are entered as the sum of k „overshoots“ with frequencies of τ years. τ is a set of integers based on the dominant frequencies in the pre-vaccination period and contains at least one seasonal term, i.e. $\tau = 1$. The term x_t is the log incidence rate of infection (Chapter 12) and cannot be observed. It describes the latent autocorrelation process defined as

$$x_{t>1} \sim \text{Normal}(\rho x_{t-1}, \sigma^2), \qquad x_1 \sim \text{Normal}(0, \sigma^2(1-\rho^2)^{-1}).$$

Das Modell wurde in einem „Bayes-Framework“ definiert, in dem die Prioritäten für die marginale Varianz und ρ definiert sind als

$$\sigma^{-2}(1-\rho^2) \sim \text{Gamma}(1, 10^{-5}), \qquad \log\left(\frac{1+\rho}{1-\rho}\right) \sim \text{Normal(0,0.15)}.$$

Quantification of vaccine efficacy

As with all drugs, vaccines require evaluation of efficacy and tolerability in randomized, controlled, clinical trials, that is, under „optimal conditions“. In many countries, a cost-benefit assessment is also required for approval. This is also to be expected for the Federal Republic of Germany in the not too distant future. Vaccine efficacy in clinical trials (VE) should not be confused with vaccine effectiveness (VE_{eff}, the effectiveness of the vaccine under everyday clinical conditions. VE is defined as the reduction in relative risk, RRR, in vaccinated compared with unvaccinated persons, of contracting the target disease. Thus, the RRR is derived from the relative risk, RR, of infection in vaccinated and unvaccinated individuals. We compare the proportion of unvaccinated persons with the proportion of vaccinated persons among those with the disease:

$$\text{VE} = 1 - \frac{\text{AR}_v}{\text{AR}_{nv}} = \frac{\text{AR}_{nv} - \text{AR}_v}{\text{AR}_{nv}},$$

$$RR = \frac{\text{AR}_v}{\text{AR}_{nv}}, \qquad \text{VE} = 1 - \text{RR} = \text{RRR}.$$

It means in this case:
AR_v : absolute risk, attack rate, of vaccinated individuals becoming infected with a pathogen against which the vaccine is intended to protect.
AR_{nv} : Absolute risk of unvaccinated individuals becoming infected with a pathogen.
RR : Relative risk
RRR : Relative risk reduction.

A vaccine must prove itself under routine conditions, i.e. independent of the „optimized“ conditions of a clinical trial. Vaccination with AstraZeneca and Biontech/Pfizer vaccines has significantly reduced hospitalizations and deaths in the United Kingdom[13]. England was the first country to start a vaccination campaign on 8 December. This April 2021 study showed that there was a 70 % (95% confidence interval 59 to 78 %) decrease in mortality and hospitalizations for COVID-19 as early as 10 to 13 days after the 1st vaccine dose. The protective effect increased to 89 % (85 to 93 %) 14 days after the 2nd dose: $\text{VE}_{\text{eff}} \approx 70 - 89\,\%$.

[13] J. Lopez Bernal, N. Andrews, C. Gower, et al, *Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study*, BMJ 373 (2021), n1088

The Reproduction Number R

Of particular importance in assessing the dynamics of an epidemic is the reproduction number R. It reflects the average number of secondary cases that an infected person can generate in a susceptible population at a given point in an epidemic. In other words, R describes how many people an infected person infects on average in a given period of time. At the beginning of an epidemic, we refer to R_0, and later to effective reproduction number, R_{eff}. Using the reproductive number R, estimates can be made of the rate of spread of a communicable disease and of the proportion of the population that must be immune after surviving the disease, by so-called „silent feiung“ (immune response of the body in the case of asymptomatic infection), or by vaccination, in order to slow or stop the further spread of the epidemic.

R cannot be used alone as a measure of the need for or effectiveness of interventions. Also important are the absolute number of new infections, the cumulative incidence, IC, or incidence density, ID, during the reference period, and the infectious dose and resulting severity of disease progression. Efforts are made to keep the absolute number of new infections low enough to allow effective contact tracing of infected persons and to avoid overloading intensive care units in hospitals. R_0 is not a globally valid constant for a pathogen, as it depends on environmental conditions and ethnic influences. Mathematically, R_0 also still depends on the underlying model (see Chapter 10).

Herd Immunity

Vaccination prevention ensures individual protection, i.e. protection of the individual and collective protection against infectious diseases transmitted from person to person, if the vaccination coverage rate in the population is high enough. This so-called herd immunity is based on the fact that unvaccinated individuals are also less likely to be exposed to the infectious agent. The proportion of individuals who must be vaccinated in a population to achieve herd immunity varies for different infectious diseases. For diphtheria, 80 % is assumed; for mumps, 90 %; and for measles, 92 to 95 %. In simple models, „effective“ herd immunity, HI, is calculated when a certain proportion of the population (HI threshold) has protective endogenous defenses. We assume that after the threshold is reached, chains of infection are broken and individuals, who cannot or do not want to be vaccinated are indirectly protected. The following applies to herd immunity

$$\text{HI} = \frac{1}{\text{VE}}\Big(1 - \frac{1}{R_0}\Big).$$

Beispiel Covid-19 und der Impfstoff BNT162b2 mRNA[14]

$$\begin{aligned}
\text{AR}_v &= \frac{8}{18.198} = 0,00044 \\
\text{AR}_{nv} &= \frac{162}{18.325} = 0,00884 \\
\text{RR} &= \frac{0,00044}{0,00884} \approx 0,05 \\
\text{VE} &= 1 - 0,05 = 0,95.
\end{aligned}$$

Thus, the vaccine was tested in a clinical trial in more than 36,000 people and achieved an efficacy of about 95 %.

R_0 was estimated to be approximately 3.0 at the onset of the SARS-Cov-19 pandemic. From this, the threshold for herd immunity can be calculated as follows:

$$\text{HI} = \frac{1}{0,95}\left(1 - \frac{1}{3}\right) \approx 0,7.$$

This result implies that approximately 70 % of the population who have not yet survived Covid-19 infectious disease would need to be vaccinated to generate sufficient herd immunity. Covid-19 recovered individuals who have developed protective immunity can be subtracted from this.

Nonspecific Protective Effects of Vaccination

The previous understanding was that a vaccine only specifically protects against a particular infection. Meanwhile, several observations suggest that live vaccines, e.g., oral poliomyelitis vaccine, BCG vaccine against tuberculosis, and measles vaccine, not only immunize the vaccinated specifically against the target diseases, but also result in protective effects against other diseases and may even prevent deaths. These broad beneficial nonspecific effects of vaccines on the immune system could be useful in a pandemic and potentially reduce the risk for a severe course of COVID-19 We have published proposals for mathematical modeling of the nonspecific effect of BCG vaccination[15].

Economic Aspects of Prevention through Vaccination

For decades, the cost of health care systems in industrialized countries has been rising faster than national income (GDP). As a result, the introduction of additional, relatively expensive vaccinations has met with opposition, despite the considerable medical benefits and cost-effectiveness that have been demonstrated, among other things, for

[14] F.P. Polack, S.J. Thomas, N. Kitchin, et al., *Safety and efficacy of the BNT162b2 mRNA Covid-19 Vaccine*, The New England Journal of Medicine 383, (2020), 2603–2615.

[15] S. Treibert, H. Brunner, M. Ehrhardt, *Compartment models for vaccine effectiveness and non-specific effects for tuberculosis*, Mathematical Biosciences and Engineering 16 (2019), 7250-4298

vaccination against *Haemophilus influenzae b*, among others, has been shown to encounter ethical and economic problems, especially when the vaccine-preventable diseases are severe but rare. These issues can also be incorporated into mathematical prediction models. Smallpox, poliomyelitis, and measles are examples of infectious diseases occurring worldwide that have been eradicated or greatly reduced in incidence by vaccination. Although there are no drugs that work well against these diseases, they have lost their terror in the population, as they are virtually absent in adequately vaccinated individuals. Since mankind has reached a high population density, pandemics caused by infectious diseases with new pathogens must also be expected in the future.

15 Mathematics Enables Cultural Events**

A simple mathematical formula can be used to determine how many people may be present at indoor cultural events.

During the Corona pandemic, as part of the social distancing, many cultural establishments such as theaters, opera houses, cinemas, etc. closed. In the context of the relaxation measures now the question arises, how many persons under which circumstances (mask, ventilation, ...) can meet again in certain in certain rooms and how cultural activities can be resumed.

A simple formula allows a calculation of the number of people that can be can be used as a guide when holding events in enclosed spaces. orientation can serve [16]. This figure can also be used as a guide for other enclosed spaces, such as schools, universities, restaurants, etc.

Prof. Dr. Martin Kriegel from the TU Berlin developed a mathematical formula for the virus variant alpha (B117) under the assumption that the (effective) reproduction number is $R < 1$[17]. The maximum possible number of persons P_{max} is calculated as follows.

$$P_{\text{max}} = \frac{\dot{V}_{\text{zu}}}{f_L \cdot f_M \cdot 105\,\frac{\text{m}^3}{\text{h}^2} \cdot t_A}.$$

The parameters on the right side are explained as follows:

$\dot{V}_{\textbf{to}}$ virus-free supply air flow rate (usually outdoor air) in m^3/h (value from the operating instructions of the ventilation system).

f_L ventilation type:

- $f_L = 1$ for Ventilation according to mixed air principle (fresh air enters the room at the top)
- $f_L = 0,7$ Ventilation according to the displacement principle (fresh air enters the room at the bottom)

f_M Factor for mask wearing:

- $f_M = 1$ nobody wears a mask
- $f_M = 0,5$ all spectators wear medical mouth-nose protection
- $f_M = 0,2$ all spectators wear FFP2 mask

t_A length of stay, i.e., the duration of the event, in hours h

The factor $105\,\text{m}^3/\text{h}^2$ is only for scaling and to give a dimensionless result for the number of people.

[16]Hygienerahmenkonzept Juni für Kultureinrichtungen im Land Berlin, Stand 7.6.2021, https://www.berlin.de/sen/kulteu/aktuelles/corona/20210607_hrk_juni_finalv3.pdf

[17]There is also a detailed web app of TU Berlin https://hri-pira.github.io/.

16 With Mathematics to Individual Therapy**

Mathematics & Pharmacy? This is a fascinating combination with exciting problems of great relevance!

Digital mobile devices are making it easier and easier to collect data, even in healthcare. But data alone is not enough to generate insights and new knowledge. You have to interpret it, understand it, and filter out the essentials from large amounts of data. And sometimes you have very little data and have to make a decision, such as in the individualization of therapies in oncology. Mathematics plays an important role here. It helps to learn from the data of previous studies and therapies to improve treatment for future patients. A report from (scientific) practice by Corinna Maier (AbbVie Deutschland GmbH & Co. KG) and Prof. Dr. Wilhelm Huisinga (Mathematical Modeling & Systems Biology; Institute of Mathematics, University of Potsdam).

Before a new drug is launched on the market, it undergoes a long research and development process. And at the end, regulatory authorities check whether it meets the high requirements for safety and efficacy. A very good overview of this can be found, for example, online at the Science Media Center Germany[18]. But how do we actually know that an approved drug is safe and effective in all (future) patients?

During the development phase, it is an important goal to investigate this question. To this end, various data on the active ingredient are collected in clinical studies, for example how much active ingredient is in the blood 1, 2, 12, 24 hours after administration. What side effects occurred? Did the expected effect occur? Are there differences between women and men? The so-called therapeutic window indicates the range in which the active substance is safe and effective (see Figure 20). But people are different, and thus the therapeutic window also varies from individual to individual. One of the many challenges in developing new drug therapies is that one does not know the individual therapeutic window. However, we do know what effects and side effects have been observed in clinical trials with test subjects and patients. From this information, one tries to form a picture. There are drugs for which the „intersection“ of all individual therapeutic windows is sufficiently large (Figure 20-C). Here, it is comparatively easy to find a single therapeutic dose for all patients. Other drugs have narrow individual therapeutic windows (Figure 20-D). Then it becomes necessary to adjust the dose for individual groups or to find an individualized dose for each patient.

The Figure 20-D also points to another important aspect, which is the main focus of this article: Even if it has been possible to find a dose that is within the therapeutic window for all, i.e. safe and effective, it is clear that for most patients it is not necessarily the optimal dose (light green dot in Figure 20-D). But how can we realize the goal of drug therapy to find safe and effective dosages for optimal individual treatment? Here, just

[18] https://www.sciencemediacenter.de/alle-angebote/fact-sheet/details/news/arzneimittel-von-der-entwicklung-bis-zur-zulassung/

in the application, an approach „takes off“, which is called *model-informed precision dosing*.

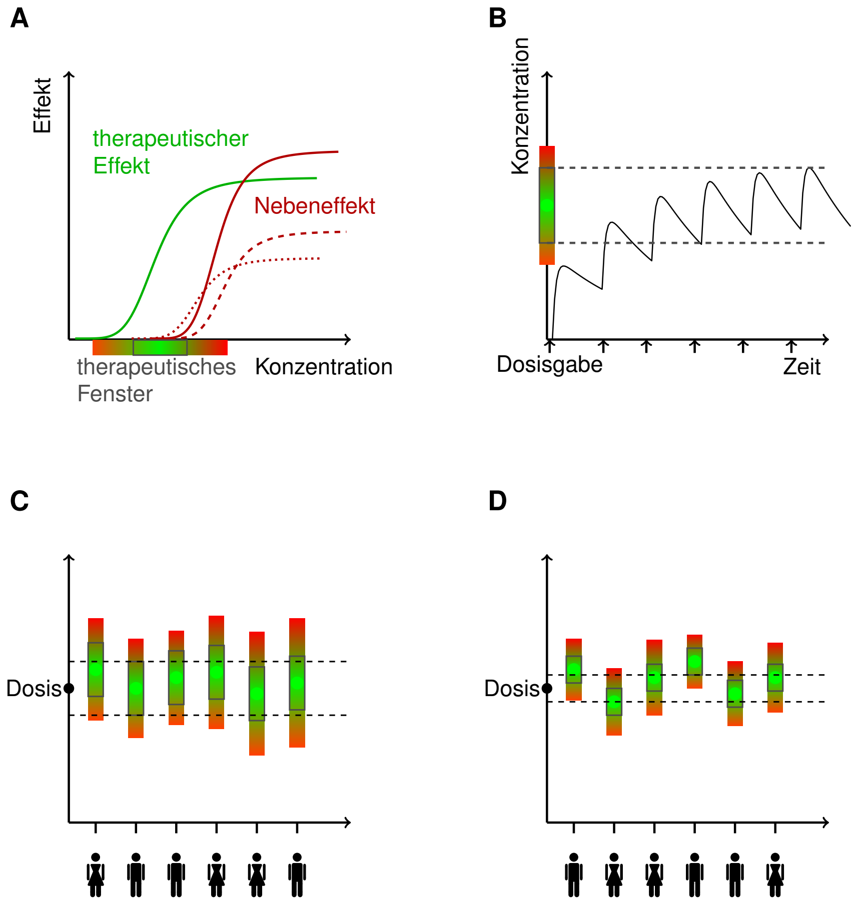


Figure 20: **The therapeutic window.** (A) Occurrence of the therapeutic effect as well as side effects as a function of drug concentration. (B) Drug concentration as a function of time for multiple dosing; in the example, concentrations remain within the therapeutic window from the fourth dose onward. Individual therapeutic windows with large (C) and small (D) intersection in the patient population.

An example from Oncology

Paclitaxel is a drug used to treat a type of lung cancer by inhibiting the cell division of cancer cells. As an undesirable side effect, it also inhibits cell division of healthy body cells. This can lead to severe side effects and the need to reduce the dose. One dose-limiting side effect is neutropenia, a drastic reduction in neutrophils (cells of the immune system) with potentially life-threatening consequences. In order to take these (side) effects into account in therapy planning, one uses models, i.e., simplified representations of reality, which depict the aspects essential for the question according to the current state of knowledge. Figure 21 shows a schematic representation of a model for a side effect of paclitaxel on the immune system. The left part of the graph describes (in a very simplified way) the processes that paclitaxel undergoes in the patient's body: Absorption by infusion, distribution in the body, and elimination. The right part of the graph describes (also very simplified) the maturation processes of blood cells in the bone marrow. The side effect of paclitaxel, the inhibitory effect on the division of stem and progenitor cells, is shown by red arrows. Since it takes a certain time (the so-called maturation time) for mature neutrophil cells to develop from the progenitor cells in the bone marrow and circulate in the blood, the side effect manifests itself in the blood only after a delay, i.e. days after the administration of the drug.

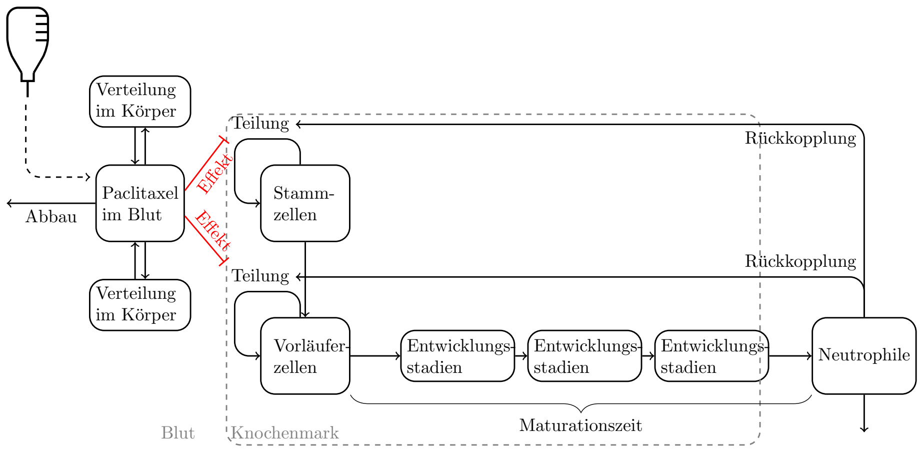


Figure 21: **Pharmacological model** for the uptake, distribution & elimination (part to the left of the red arrows) as well as the effect of the active substance (part to the right of the red arrows). Such a model can be translated into mathematical equations: each box corresponds to a variable, each arrow to a term in the underlying mathematical equations. The graphical representation allows pharmacists and mathematicians to talk about the model on a common basis.

The Bridge to Mathematics

Mathematically, one can translate such representations into functional equations. These depend on the patient's condition, e.g. the concentration of stem cells and progenitor cells in the bone marrow and neutrophil cells in the blood. And on certain parameters, let's call them ϑ (a little Greek theta), such as the maturation time or how effectively the drug inhibits cell growth. Assuming for a moment that we had all of this information, we could use the model to predict, based on the patient's current state x_0 at time t_0 (time of dosing) and the individual parameters ϑ, the patient's state $x(t_0 + h)$ is expected to decline after h days:

$$x(t_0 + h) = f(h, x_0; \vartheta, D) + \epsilon_k \,.$$

Where f is the model prediction function and D is the drug dose, that was administered. The term ϵ_k is also important; it expresses inaccuracies in our model, e.g. due to model errors or measurement inaccuracies, etc. So by ϵ_k we cannot make accurate predictions, but have to live with some uncertain prediction. Just as one would expect given the complexity of the problem. The predicted patient state $x(t_0 + h)$ includes all aspects of the model and can be used to calculate further information, e.g., how low the neutrophil concentration N_{12} has fallen in the blood after $h = 12$ days:

$$N_{12} = g\big(x(t_0 + 12)\big); \qquad t_0 \mathrel{\widehat{=}} \text{time of last dosage}$$

Here, one uses another function g for information extraction. Thus, we can predict, to some residual uncertainty, how low the concentration of neutrophils in the blood will fall when the drug dose D is administered. Of course, we will also model and consider the effect on tumor growth at the same time. This aspect is not shown in the above model for simplicity. The difficulty now is to find the right dose to, on the one hand, prevent the cancer cells from proliferating and, on the other hand, ensure, that the neutrophils are not reduced too much, so that the patient is still sufficiently protected against infections.

The previous story has only one small catch: In clinical reality, we do not know the patient state x_0 and its individual parameters ϑ. Here again mathematics, or more precisely statistics, can help. We do not yet know how the drug works in our patient, but we do know how it has worked in many other patients. We can use this information from previous clinical trials to estimate a probability distribution $p(x_0, \vartheta)$ on the states and parameter values. This indicates how often a certain state or parameter value has been observed so far. At the beginning of therapy t_0, we assume that each state and parameter value is as likely to describe the state of our patient as $p(x_0, \vartheta)$ indicates. We can also say that $p(x_0, \vartheta)$ expresses our knowledge/unknowledge about the initial patient state x_0 and the individual parameters ϑ. At the beginning of therapy, this already allows model predictions of what course of neutrophil concentration we expect based on certain patient characteristics (e.g., age, weight). Individual patient data collected during the course of therapy, e.g. neutrophil concentration by blood samples, are then used to assess our patient's condition more and more accurately.

Model-Assisted Dose Individualization

Now, how can the physician make an informed dose decision based on these two different sources (a priori model prediction, individual patient data)? Now this is where mathematics comes into play again: In Bayesian statistics, one uses the following formula:

$$\underbrace{p(\vartheta, x_0|y)}_{\text{knowledge update}} = \frac{\overbrace{p(y|x_0,\vartheta)}^{\text{indiv. data}} \cdot \overbrace{p(x_0,\vartheta)}^{\text{prior knowledge}}}{\underbrace{p(y)}_{\text{normalization}}},$$

to link the general prior knowledge of many previous patients with individual patient data y and update the knowledge about our patient. The branch of mathematics that deals with linking models and data is called data assimilation. Data assimilation techniques have been developed in meteorology (weather forecasting) as well as geology. Similarly, for individual therapy planning, model prediction can be improved with each measured neutrophil concentration of the patient and different scenarios for possible future dose administration can be generated. Figure 22-A shows the measured neutrophil concentrations (crosses). The green area shows the simulation (incl. remaining uncertainty) based on the updated knowledge. The small black arrows indicate when the drug was dosed: thus, the right part without crosses is a prediction into the future (based on the bishereditary dose). The red areas indicate the degree of adverse reaction; the darker the red, the more severe the degree. The degree of side effect refers to the lowest neutrophil concentration after drug administration. As mentioned earlier, all predictions are subject to uncertainty. Figure 22-C shows, among other things, the probability of expecting grade 0-4 side effects (i.e. with 60 % grade 3). One can also predict when the patient will have recovered from the side effects, i.e. when the neutrophil concentration will be close to baseline again. This is of great interest for therapy planning in a hospital, because the next dose can only be administered when the levels are high enough again.

The doctor has to take many different aspects into account when planning the therapy (general condition of the patient, change in tumor size, other side effects, etc). Regarding neutropenia, she would choose the dose so that neither grade 4 occurs (since life-threatening), nor grade 0 (since then the drug dose would be too low to sufficiently inhibit tumor growth). In Figure 22-B, the model was used to predict the probability of grade 0 and grade 4 (light and dark red) and grades 1-3 (green and yellow) occurring for four different doses. The two horizontal dashed black lines are to indicate a target range: Grade 0 with less than 10 % probability and Grade 4 with less than 5 % probability. Thus, among the predictions of the four doses shown, one would select the second from the left according to these criteria, i.e. reduce the previous dose by 15 %. Such predictions provide evidence-based decision support for the clinician because they integrate information from the different sources (clinical trials, patient data, literature).

And there are many more exciting & important problems to solve: How to derive general dosing recommendations from the totality of individual doses? How can knowledge from therapy individualization of previous patients be used to individualize therapy even better & faster in the future, without having to share personal patient data? Such research

[19] Image based on article by C. Maier et al, *Bayesian Data Assimilation to Support Informed Decision Making in Individualized Chemotherapy* (CPT:PSP), 2020

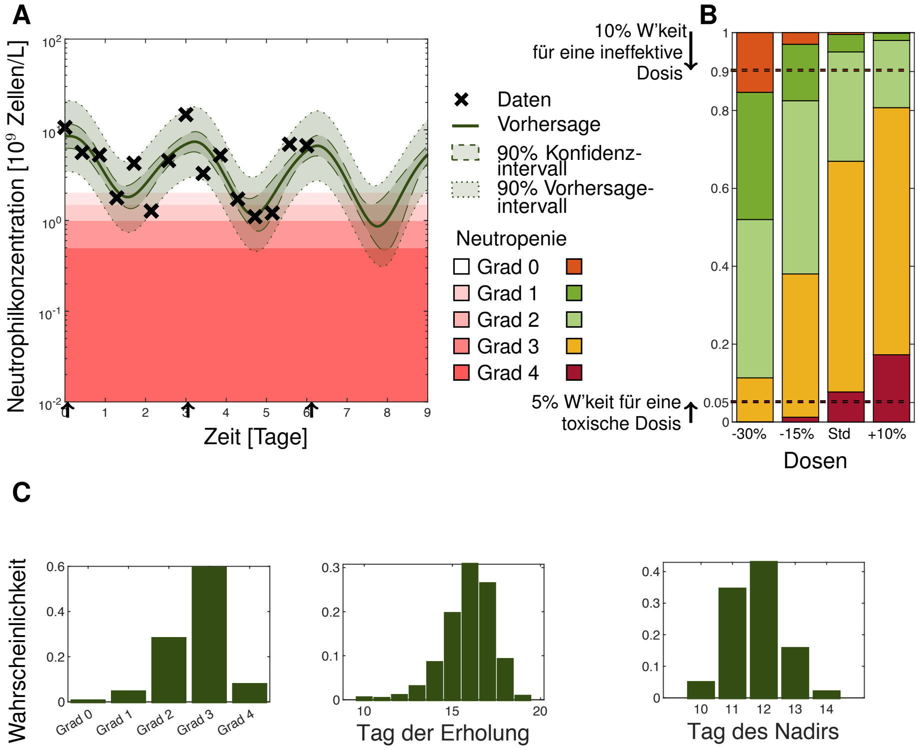


Figure 22: **Aiding decision making for optimal individual dose.** (A) Individual data (crosses) and model prediction as a function of time. (B) Prediction of neutropenia levels for previous dose and alternatives (2x reduced dose, 1x increased dose). (C) Model-predicted probability of neutropenia grades, day of recovery, and day with lowest neutrophil concentration (see also Eq. with N_{12} earlier in the text). [19]

questions are being investigated, for example, within the PharMetrX: Pharmacometrics & Computational Disease Modelling Berlin/Potsdam doctoral program [20] and the DFG Collaborative Research Center on Data Assimilation[21] at the University of Potsdam. In the pharmaceutical industry, special pharmacometrics departments deal with the analysis of data from clinical trials to support decision making in the drug development process.

Those interested in mathematics with pharmacy application should choose an undergraduate mathematics major, e.g. with a minor in biology, or pharmacy if possible. A solid basic education is important; you can specialize later. Most important, however, is motivation and enthusiasm for mathematics and a genuine interest in its application.

[20] http://www.phametrx.de
[21] http://www.sfb1294.de

17 Smarter Planning makes Healthcare more Efficient and Safer*

The healthcare sector faces substantial challenges such as high work pressure, long waiting times and high costs. However, better planning based on probabilities can solve these problems.

In the next two decades, healthcare costs are estimated to rise from ten to twenty percent of our gross national product; an enormous increase in expenditure. The surgical area in hospitals is particularly expensive, accounting for more than 40 % of all hospital costs. One reason for this is the complexity of operations, where the right resource must be available in the right quantity at the right time for the right patient. The shortage of staff often leads to situations where operating rooms are left vacant, further burdening the hospital: Accumulated delays during the day force overtime. This not only results in costs for the hospital, but also creates stressful situations for staff and patients and has a negative impact on the quality of care.

OP-timization

Nevertheless, Prof. Ralf Borndörfer (FU Berlin & ZIB, Zuse Institute Berlin), Dr. Guillaume Sagnol (TU Berlin) and Alexander Tesch (ZIB) are convinced, that mathematics can help solve this problem by making better use of human and material resources. They conducted a project in collaboration with the Charité University Hospital in Berlin, Germany, in which an algorithm was developed to create optimized operating (OP) room schedules based on individual target criteria and constraints; see Figure 23. The goal is to more efficiently allocate, sequence, and control OP processes with respect to given human and material resources. But what can an algorithm do better than an experienced OP planner? Let's consider an example: Let's assume that ten surgeries have to be distributed and sequenced across three operating rooms for tomorrow. For this small scenario, there are already about 250 billion different OP plans. With 55 surgeries and 15 OP rooms, there are about. $1.96 \cdot 10^{87}$ different plans, which is about the number of atoms in the universe. And this does not even include the uncertainty in the operating room durations.

Accordingly, there are far more possible surgical plans than humans can evaluate. An algorithm is much faster, even if testing all possible surgical plans would take years, even on a supercomputer. And this is where mathematics comes into play: clever methods can be used to narrow down the search considerably, so that, in the end, only a few thousand surgical plans may need to be evaluated. On a standard PC today, this can be done in seconds or minutes, depending on the size of the problem. The pure time required to create an OP plan can thus be significantly reduced.

Schematically, we can think of these scheduling problems as a Tetris game (see Figure 24), where the shape of the pieces represents the need for different resources over

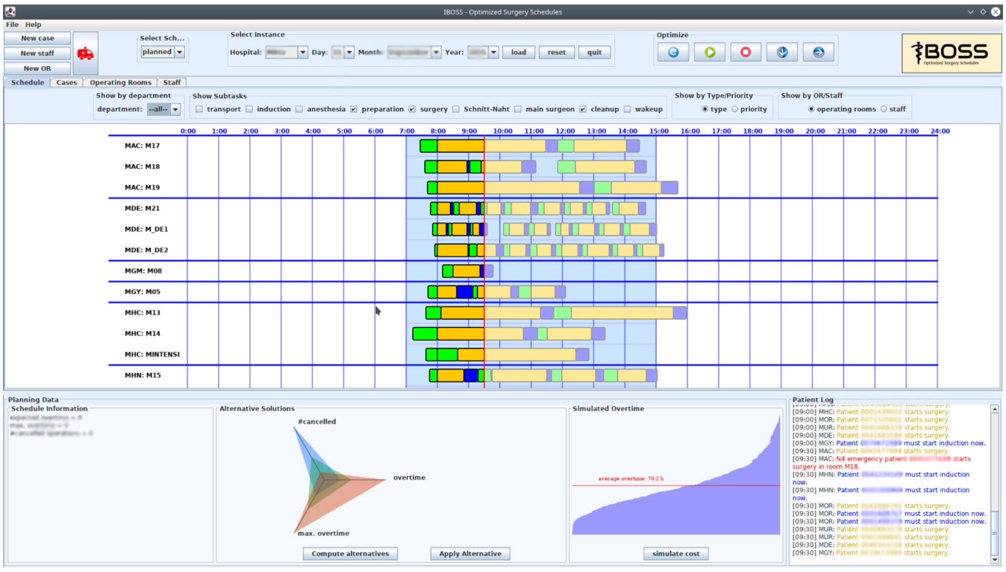


Figure 23: Prototype of the developed planning software.

time. The goal is to fit the pieces together as best as possible, according to criteria that can be selected and adjusted (number of overtime hours, staff idle time, patient waiting time, schedule stability,...). Now the question is what an „optimal“ OP schedule looks like. Each person involved probably has a different idea of the ideal schedule: The scheduler wants to make the best use of OP resources, the staff wants to finish on time, and the patient expects the best possible quality of care. In practice, however, there will be no OP schedule that maximally satisfies everyone's wishes, as the target criteria are often in conflict. Algorithms allow surgery plans to be optimized with respect to a desired weighting of the target criteria. This allows OP plans to be generated that represent the best possible trade-offs between multiple target criteria. If the calculated OP plan does not correspond to the desired result, the weighting can be adjusted until a satisfactory solution is obtained.

Randomness and Reactivity

Uncertain events have a strong impact on surgical planning. Rarely are OP plans executed exactly as planned the day before. One of the main reasons is the often high variance in OP durations, which makes accurate planning difficult. Mathematics helps to predict OP durations. This requires one thing above all: data, the more the better. If the data set is large enough, a probability distribution can be calculated for each OP duration. A probability distribution is much more meaningful than a fixed average value, since uncertainty occurs to varying degrees. That is, an OP plan that looks good on the page with respect to fixed OP durations might not be feasible in reality. By frequently simulating the OP day, algorithms can better quantify the existing uncertainty in

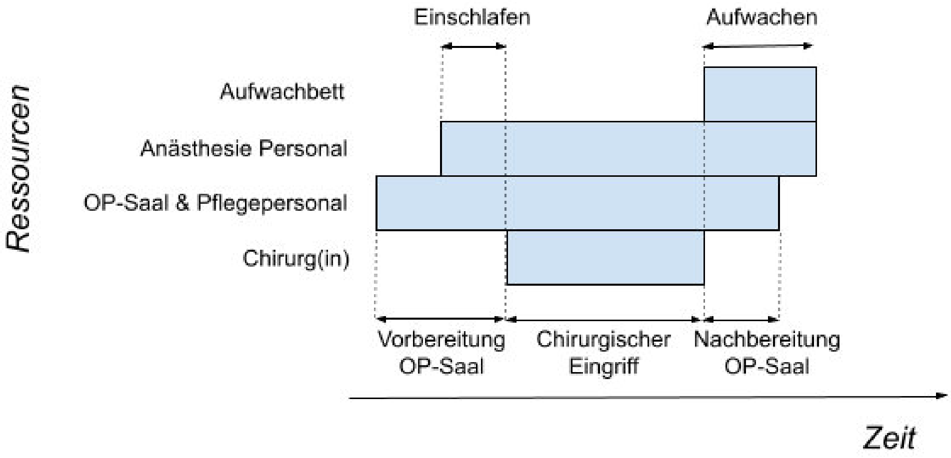


Figure 24: Tetris representation of a surgery.

the system. This in turn allows better planning decisions to be made, similar to the risk management of a good game of poker.

An important issue in dealing with uncertainty is to find reactive strategies that adapt to the observed information. For example, an optimal strategy might be as follows: „If the OP in room A is completed first, perform OP X in room A; otherwise, perform OP X in room B.“ However, such adaptability is not always possible in a hospital setting, e.g. to ensure surgeon-patient bonding. This gain from adaptations can also be studied theoretically. For example, Dr. Sagnol's team proved that, given a criterion with fixed costs per operating room and variable costs for overtime, there exists a simple non-adaptive strategy that loses at most $100 \cdot e^{-1}\,\% \approx 36\,\%$ compared to an optimal adaptive strategy. On the other hand, if the goal is to finish all operations as early as possible, then any strategy that considers rescheduling with some delay can be arbitrarily bad. However, the loss of optimality grows very slowly (and doubly logarithmically) with the number of operating rooms.

18 Better Images*

New calculation methods make medical scans faster and more accurate. They require fewer photos to produce a good image. In this way, these mathematical methods help in the study of bone diseases.

The population is aging and bone diseases are becoming more common. For example, one in three women over the age of sixty suffers from osteoporosis, a condition in which the bone slowly disintegrates. Researchers are trying to figure out exactly what happens to bone in this disease. They want to scan the same bone over and over again, but with all these scans, the amount of radiation on the body becomes unacceptably high. Mathematician Joost Batenburg, a researcher at the Center for Mathematics and Computer Science and a professor in Leiden and Antwerp, is therefore developing methods that use less radiation: „In small studies, it has now been possible to reduce the radiation dose by a factor of three. "

Work with Prior Knowledge

Dutch hospitals perform more than one million CT scans each year. A CT scanner takes X-ray images from different angles, all of loose body parts. Then the trick is to reconstruct the full three-dimensional image from these two-dimensional photographs.

The body is made up of all sorts of tissues, each with its own density, all of which are given a different gray color in the scan. But in discrete tomography, Joost Batenburg's area of expertise, the assumption is, that there are only a limited number of different densities: „This works well with bone, for example. It has about the same density everywhere and is very porous on a small scale. You see an object with many holes on the scans and you only need two colors: there is bone or no bone."

By using this prior knowledge about the structure of bones, the total number of possible solutions decreases dramatically: Instead of all sorts of shades of gray, you now use only black and white. This makes it much easier to find the right solution from all the possibilities. In this way, it is possible to calculate the correct three-dimensional image from fewer photos. The patient is in the scanner for a shorter time and receives less harmful radiation.

Although it is very useful to be able to work with a small amount of measurement data, the problem with only two densities seems to be mathematically more difficult to solve. Batenburg: „In the standard problem, the solution can have all possible shades of gray. It's a continuous problem, and you can differentiate, integrate, whatever you want. As soon as you start dealing with discrete problems, you lose all these nice analyses." This adds significantly to the computation time. To give you an idea of the scale of these problems: A scan quickly consists of 2000 x 2000 x 2000 elements. This results in a system of eight billion comparisons with eight billion unknowns. To really calculate faster with this, a deep mathematical understanding is required.

The Golden Mean

Other medical images have exactly the same problems. In science, powerful algorithms have been developed in recent years that can process little data but require a lot of computing time. This makes them useless in practice. With these methods, a patient can get out of the scanner faster, but would then have to wait days for the results because the computer is still calculating.

The old-fashioned methods are fast, but they need many more images to do their job. Joost Batenburg is working with graduate students to find a happy medium: He's adapting the old computational methods to combine their speed with the new technique, which gets good results with fewer photos. „The standard algorithm from the seventies is a *filtered back projection* and contains, as the name suggests, a filter. We only adjust this filter and leave the rest of the algorithm unchanged. This makes it easy to change code that is already in all sorts of devices." With this solution, a whole set of methods that were initially only scientifically useful can be applied in practice. In this way, the calculation can be performed up to a hundred times faster.

The new methods are not yet used in practice. When doctors hear about them, they ask if everything has already been clinically tested. Without these tests, they are not allowed to use the technique. The medical world is conservative, and change is slow. It takes many years for innovations to be used in practice. Batenburg: „You have to give it a lot of time, but I expect hospitals to be using our methods within ten years. "

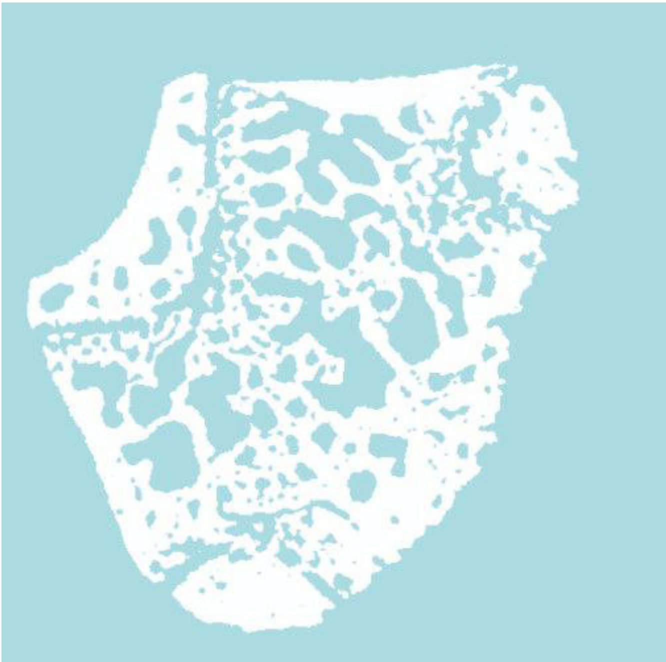


Figure 25: This section of the femur (of a mouse) is the result of a discrete tomographic reconstruction based on data from the scan.

19 Dieting is a Matter of the Mind*

Appetite for food decreases weeks before starting a diet, but also increases again weeks before the end of the diet. This is the result of the first mathematical model of the eating sense.

When man lived as a hunter-gatherer, it was wise to eat his fill, when there was just there was an abundance of food. Scarcity was the rule then, abundance the exception. But now that Western man has a constant abundance of food, it is very unwise to eat this abundance every time, no matter how tasty our appetite is. For many people, this is difficult, considering that in Germany, about a quarter of adults (23 % of men and 24 % of women) are severely overweight (obese)[22].

The list of diets to combat all this excess weight goes into the dozens: From Atkins to Montignac and Sonja Bakker to saltless diets or interval fasting. People on diets often keep a record of it like an accountant, how many calories they consume each day and, often, how much they burn through daily exercise and sports.

Jojo-Curves

It occurred to mathematician Johan Grasman of Wageningen University that people keep records, but actually have no idea how their appetite for food evolves over time. „Whether we want to eat or not is partly a physical process, but also partly a mental process, a conscious decision,„ Grasman says. „I wanted to develop a formula to describe the desire to eat. If we know how the desire to eat changes, then we might be better able to respond to how a person can best line up.“

People who gain weight take in more calories than they burn. The difference is stored in fat tissue. The fat cells produce the hormone leptin. When the body has produced

[22] https://www.rki.de/DE/Content/Gesundheitsmonitoring/Themen/Uebergewicht_Adipositas/Uebergewicht_Adipositas_node.html

enough fat, leptin gives the signal to eat less. Leptin is an important hormone, but not the only one that contributes to the desire to eat.

The crucial new factor in the model is the „urge“. Under this term, Grasman subsumes all stimuli (except leptin) that lead to, someone to eat or, in some cases, not eat. These can be physiological stimuli, that cause the body to eat (e.g. insulin), or psychological stimuli such as the desire to lose weight.

Grasman modeled the so-called 'lines' as a system of *differential equations*, which describe the desire to eat („urge“), calorie intake, and calorie consumption as a function of time. An (ordinary) differential equation contains unknown functions and their derivatives, usually with respect to time.

The variables in his model are the amount of calories stored in adipose tissue and the level of plasma leptin. The model has as input a drive that controls food intake. This drive consists of a collection of physiological and psychological incentives to eat or to stop eating.

To estimate some of the parameters in his model, he used data from the scientific literature from two women, who tried to reduce their weight five times over a ten-year period. The result was the familiar and unhealthy yo-yo curves: Weight loss – gain – loss – gain in body weight over that ten-year period. Subsequently, Grasman was able to use his formula to calculate the desire to eat over the same ten-year period. After a period of weight loss, dieters often fall back into old eating habits. With Grasman's model, one could better understand how how a person's urge to eat works, and possibly derive new strategies from it.

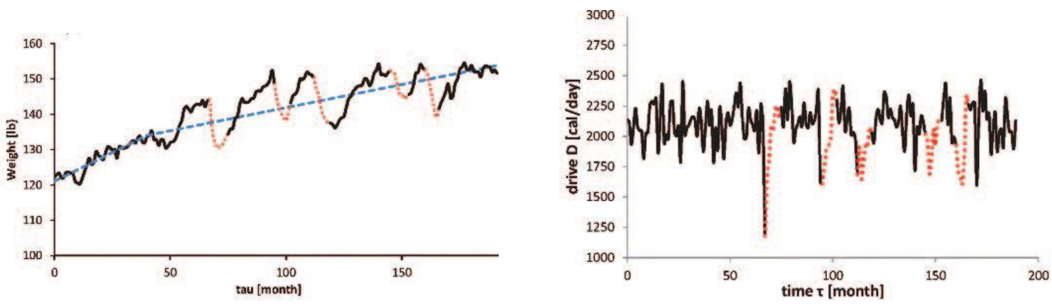


Figure 26: Two diagrams on weight and 'desire to eat'. The first shows weight as a function of time. The second shows the desire to eat as a function of time. The red dots are the data in the conscious attempts to lose weight.

Lose Weight Starts in the Head

The conclusions turned out to be surprising. Grasman: „First of all, the appetite for food seemed to diminish weeks before the diet began. Apparently, these women had been thinking about it for some time. Second, it turns out that the appetite picks up rapidly weeks and sometimes even months before the diet ends. Apparently, the mind has become weak again. These two conclusions show how important the mind is. Just think about 'comfort eating': If one day you don't like it, you can make eat a sweet to make you feel better."

According to Grasman, the psychological processes involved in the lines, much greater than the role of physical processes: „The hormone leptin reduces the desire to eat by only a few percent. Diet reduces the desire to eat by about 10 %. But the rest, more than 80 % of the desire to eat, has to do with what the mind wants."

Traditionally, physiologists left out the mind, when they looked at the lines, and psychologists left out physiology. Grasman is the first to model body and mind simultaneously. He already has ideas about how to break down the eating sense further, so he can better show the contribution of different hormones, how carbohydrates and fats separately contribute to weight gain, and perhaps even disentangle psychological processes.

20 The Optimal Dose Finding of Antibiotics**

The correct dosage of antibiotics plays an important role in the treatment of infectious bacterial diseases, as both too low and too high a dosage can negatively affect the outcome of treatment.

In his bachelor thesis at the University of Wuppertal, Niels Neveling investigated how to mathematically determine the optimal dose finding of antibiotics[23]. How often should one administer which dose of an active substance?

A dose that is too low may not completely eliminate the pathogen and thus may lead to the selection of resistant bacteria. On the other hand, a dose that is too high may not provide an improved result or may even harm the macro-organism (e.g., humans), as some side effects are dose-dependent. Therefore, it is important to support optimal dose finding with a mathematical model and optimize the effectiveness of dosing.

First, pharmacokinetics (PK) describes the time-dependent distribution of the drug, i.e., the antimicrobial substance, in the organism. Important indices can then be derived from this, namely. the maximum concentration (C_{max}), the area under the serum concentration-time curve (AUC), i.e. the loading of the drug in a given time interval and the half-life $T_{1/2}$. The first two indices depend on the daily dose and the dose interval, the latter depending on the drug and the organism to which the drug is applied.

Pharmacodynamics (PD) describes the effect of the active ingredient, e.g. the rate of killing of the pathogen, compared to its concentration in the body. The minimum inhibitory concentration (MIC) indicates, from which concentration the substance (so-called. bacteriostatic) inhibits any growth of the bacterial population. Although the MIC value describes the potency of a substance, it does not provide any information about its time-dependent activity. The time dependence of the effect of an of an antibiotic can be determined by taking multiple samples, and the resulting obtained bacterial count reduction curves are expressed as „time-kill curves“.

These so-called PK/PD parameters C_{max}, AUC, MIC relate the antibacterial effect to the time of exposure and quantify the activity of the antibiotic. The C_{max}/MIC rate gives the ratio of the highest concentration to the minimum inhibitory concentration. T>MIC is the percentage of a dosing interval in which the concentration of the drug is above the MIC. Finally, the AUC/MIC (or 24-AUC/MIC) indicates the ratio of the amount of of drug to which the bacterium is exposed over 24 hours and the MIC. is given. Using differential equations, one can model the relationship between the concentration of the antibiotic and the effect that occurs, and with integration over a dosing interval, one gets the dosing intensity D. This dosing intensity D is an indicator of the potency of an antibiotic against a pathogen and allows conclusions to be drawn about the killing properties of the antibiotic.

[23] Niels Neveling, *Mathematische Modellierung von Pharmakokinetik und Pharmakodynamik antibiotischer Substanzen*, bachelor thesis, University of Wuppertal, 2015

The following differential equation can be derived from the interaction of the active substance with the organism via so-called receptors. This relates the concentration of the active ingredient and the concentration of the active ingredient-receptor compound (only in this compound is an effect on the pathogen possible). Association describes the connection of γ drug molecules with a receptor and dissociation the corresponding dissolution of the connection.

$$\dot{v}(t) = \overbrace{k_{+1} \cdot c^{\gamma}(t) \cdot [r_0 - v(t)]}^{\text{Association}} - \overbrace{k_{-1} \cdot v(t)}^{\text{Dissociation}} \quad \text{with} \quad v(0) = 0.$$

Assuming that the above differential equation reaches equilibrium in the shortest possible time, the concentration of the drug-receptor complex can be determined as a function of the drug concentration. The dosing intensity D is then the mean value of the concentration of the drug-receptor complex in a certain time interval τ.

$$D = \frac{1}{\tau} \int_0^{\tau} \frac{k_k \cdot c(t)^{\gamma}}{c_{50}^{\gamma} + c(t)^{\gamma}} \, dt.$$

The three parameters (K_k, C_{50}, γ) describe the pharmacodynamic profile of a drug and can be determined from experimental data.

To calculate this dosage intensity D, it is necessary to model the degradation of the antibiotic in the body using a multi-compartment model (similar to the SIR model from chapter 10). Here, the compartments represent the parts of the body outside the bloodstream that the drug reaches and is degraded there under certain assumptions. By means of equalization calculation (see Chapter 12), the functions are parameterized on the basis of given measurement data (obtained by blood tests). Figure 27 shows an example of the concentration curve for the three dosing intervals $\tau \in \{24, 12, 6\}$. Figure 28 shows the analogous result for a 2-compartment model. Figure 29 shows the curves of the dosing intensities to the three profiles, again recording either the dosing interval or the daily dose.

As part of the bachelor's thesis, the effect of the antibiotic ciprofloxacin on populations of pseudomonads (causing, among other things, pneumonia) of different sensitivity was investigated in the test tube. For this purpose, dosing intensities were determined at different initial concentrations. A dosage of ten times the MIC value of the respective population was found to be efficient, regardless of the sensitivity of the pathogen population.

The PK/PD parameter relevant for a therapeutic effect is the ratio AUC/MIC as well as C_{max}/MIC. However, these parameters can naturally only be measured after application of the substance as a tablet or injection, but can certainly be simulated in a test tube experiment. In the experiments on which this bachelor thesis is based, the indicator strains, which had different MIC values, were exposed to constant compound concentrations. As optimal for achieving a good therapeutic outcome, i.e., clinical cure of the patient and avoidance of resistance development, is considered to be a C_{max}/MIC ratio of 10. This theoretically necessary dose increase must then be balanced with the side effect profile of the substance and must be consistent with the approved doses.

Calculations such as these can indicate a necessary increase in dose, if necessary, and thus help to help to individualize therapy by determining the optimal concentration for killing germs. Likewise, therapy is economized because the dose can be individualized and excessive doses can be avoided. Furthermore, ineffective doses can be identified because a C_{max}/MIC rate necessary for effective therapy cannot be achieved with approved dose regimens. Finally, therapy failure can be prevented and resistance development of the bacterial pathogen due to too low exposure can be reduced.

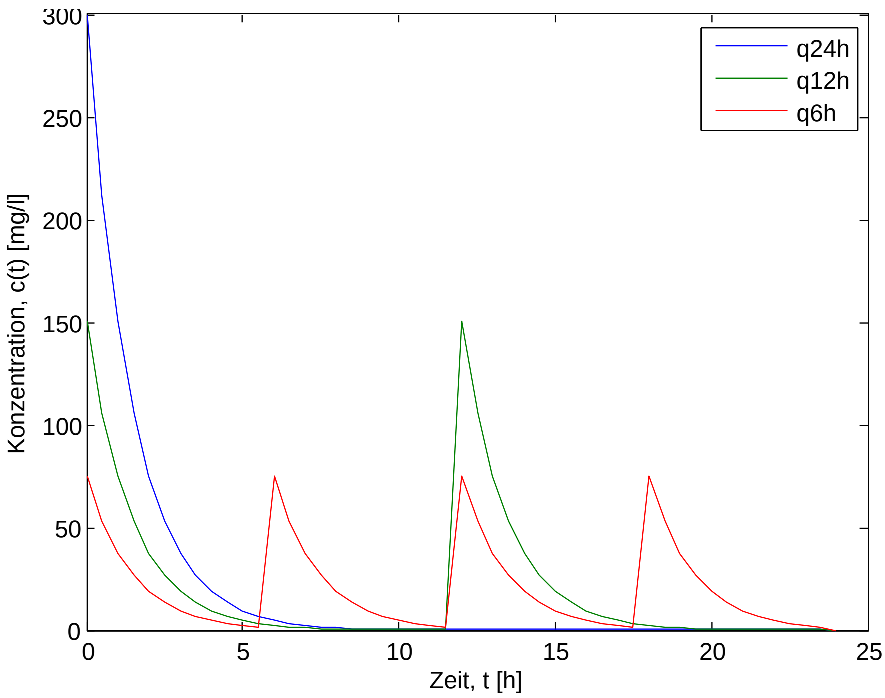


Figure 27: Pharmacokinetics in the 1-compartment model.

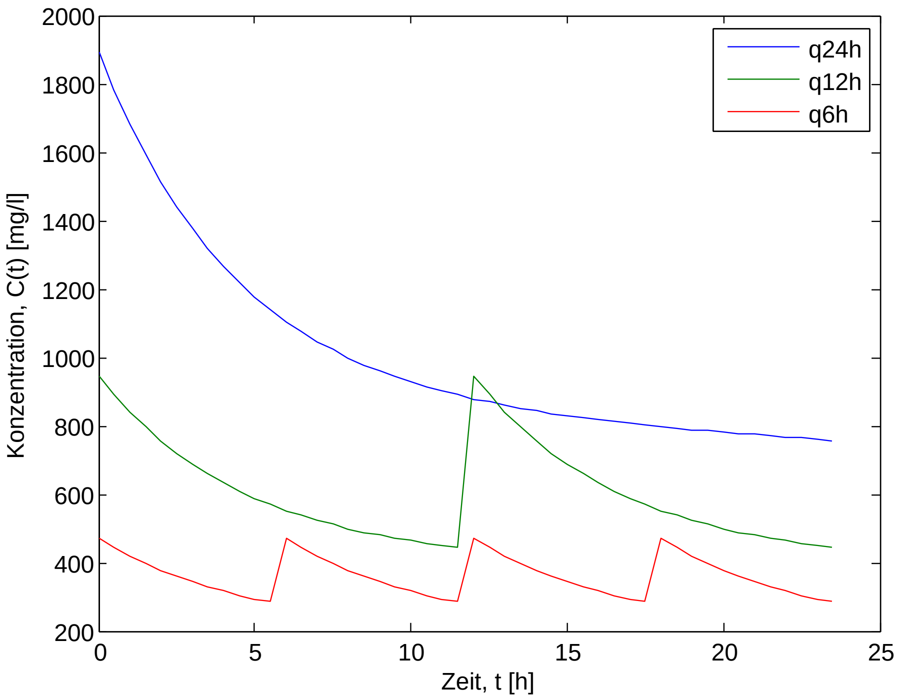


Figure 28: Pharmacokinetics in the 1-compartment model.

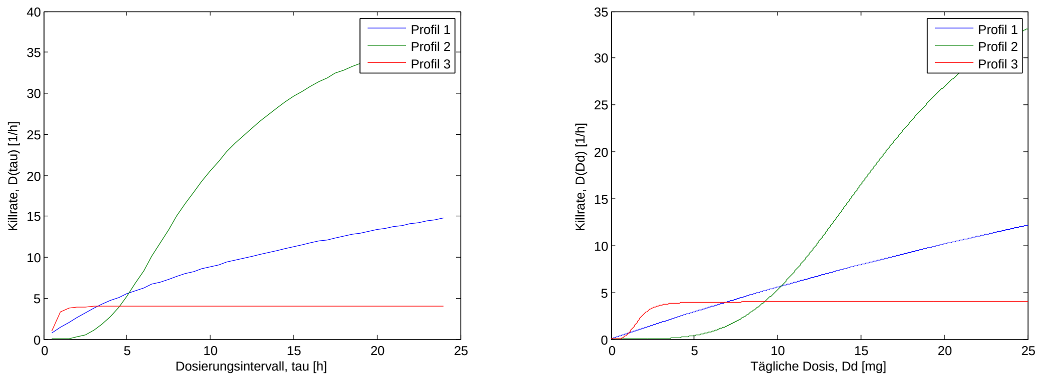


Figure 29: Left: 10mg daily dose (Dd), Right: 5h dosing interval (τ).

21 Mathematics in Chemistry: the Search for Synthetic Pathways for New Drugs*

Not only viewers of the series *Breaking Bad* but also many scientists consider chemical synthesis to be a kind of art form. Finding good chemical synthesis routes is a major challenge in the search for new active ingredients in medicine or crop protection. The fact that mathematics and artificial intelligence can help here may come as a surprise.

Dr. Georg Mogk, Principal Expert Applied Mathematics, Bayer AG, Leverkusen and Priv. Doz. Dr. Thomas Mrziglod, Head of Applied Mathematics, Bayer AG, Leverkusen give an example of the role of mathematics in chemistry.

Mathematics has the reputation of being very abstract and perhaps also somewhat unworldly. In the following, we will show that this is not the case, using chemistry as an example. Here, mathematics can help to bring conceivable molecules (e.g., as active ingredients of drugs) into reality. The image of the chemist tinkering around in his laboratory and finding sensational new active ingredients is increasingly becoming a thing of the past. In fact, potential active ingredients are designed on the computer for a specific effect. Algorithms are used to generate hundreds of thousands of substances that could have a specific effect. For the most promising candidates, this must then of course be tested in the laboratory. To do this, each drug candidate must be synthesized in sufficient quantities.

Figure 30: Synthesis robot at Bayer

However, predicting whether a molecule can actually be synthesized is very difficult. Experts estimate that for every synthesizable molecule, there are at least a billion molecules that could exist but cannot be synthesized with the current state of chemistry. This leads us to two questions: *is it possible to predict whether a computer-designed molecule is synthesizable?* And if so, *how can it be synthesized?* The concept that can be used to answer these questions is called retro-synthesis and was developed starting in the late 1960s by American chemist Elias James Corey.

In retro-synthesis, one mentally goes through each bond of a molecule, separates them and asks oneself through which reaction scheme the bond could be formed. Further, one asks what the starting products (also called reactants) would have to look like in order for a forward synthesis to produce exactly the bond one is looking for. Recursively, one now applies the same procedure to the reactants just determined, until one finally ends up with purchasable molecules or in a dead end. The result is then the so-called synthesis tree, which contains all conceivable synthesis routes. The chemist must then *only* select the best synthesis route from the synthesis tree. Corey was awarded the Nobel Prize in Chemistry in 1990 for the concept of retro synthesis.

As simple as the concept sounds, it has its pitfalls, which can be summarized in two key questions:

- Where and how can you cut a molecule in the first place?
- The synthesis trees can get very large. Is there a synthesis pathway that ends up exclusively with purchasable building blocks?

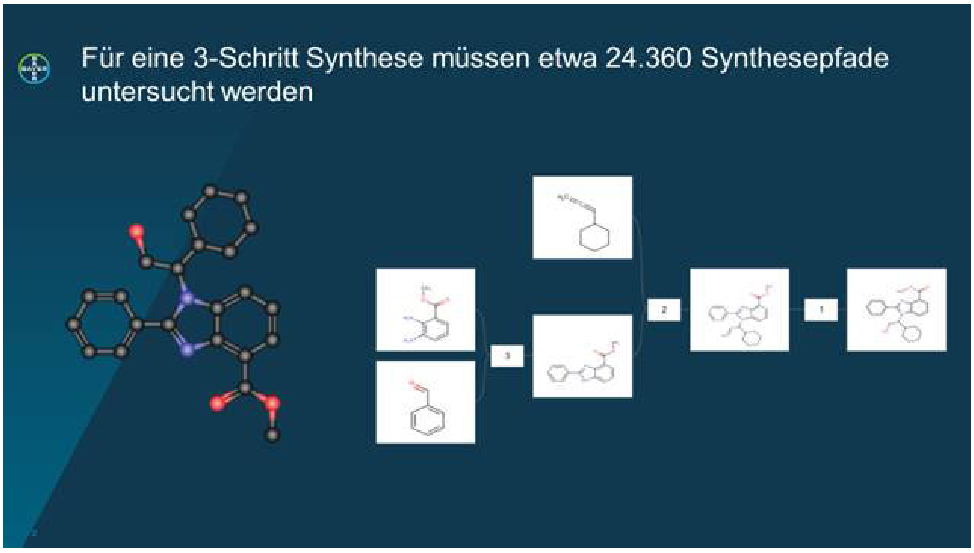


Figure 31: Three-step synthesis route

This is where mathematics comes into play. Since there are extremely many possibilities for both questions, computers lend themselves to addressing them. Rule-based approaches for the first question quickly reach their limits due to the exploratory chemical knowledge. A more modern approach is to let the computer learn the tailoring rules itself. Updating the system as the chemistry evolves is then very simple: let the computer learn on the new data. We are pursuing this approach within Bayer AG. To do this, we have used 19 million published reactions from the last 250 years and our own laboratory journals, to train an artificial neural network.

Before the computer can learn, however, there are some obstacles to overcome: The chemical nomenclature for molecules and reactions is very accessible to humans, but it must first be converted into a representation of molecules that is suitable for computers. There are also challenges with the reactions that need to be learned: In publications, researchers focus on the essentials. „Unimportant“ byproducts, such as water or CO_2 are often not documented. Or atoms from the solvent suddenly play a role in the reactions, but do not appear in the reaction equations. In short: in many published reactions the mass balances are not correct. This causes great problems for computers to understand reactions correctly and one has to help out with intelligent mathematical algorithms (keyword *atom mapping*).

Another challenge lies in publication behavior, as usually only responses that work are published. However, when we humans learn, we also learn from what doesn't work. This is also true for computers: „humans learn from their own mistakes, computers from the mistakes of others“. The main point is to find *meaningful* reactions that do not work. These are those where it is not immediately obvious to an expert that the reaction does not work. We find such reactions in Bayer's laboratory journals, for example. Bayer is over 150 years old, and chemical research at Bayer has produced some 12 million electronic laboratory journal entries in recent years. However, these must be read and interpreted. We also use a neural network to automatically interpret the laboratory journal entries. The neural network reads all 12 million lab journal entries and evaluates each one whether the reaction described there worked or not.

After all these preparations, we have a machine-readable data set of functioning and non-functioning reactions. With this data, we train a neural network for a so-called one-step retrosynthesis.

This would give us a modern answer to the first of the two questions above, *where do I cut?*. Let us now turn to the second question: *How do we find a synthetic pathway that ends up with known or purchasable starting materials?* In answering this question, we struggle with what is called *combinatorial explosion*. Even though we are dealing here with the synthesis of so-called *small molecules*, these can have up to 100 connections between heavy atoms. For example, let's take a molecule with 100 compounds, which we want to synthesize in 10 steps. Then we would have to investigate $5 \cdot 10^{21}$ possible synthesis routes. For comparison, the universe has only existed for about $7.8 \cdot 10^{18}$ seconds. But are 10 steps enough?

To cope with this enormous variety, we apply algorithms developed for strategy games, such as chess or GO. As in strategy games, it is generally not possible to explore the entire tree. Therefore, efficient algorithms are needed to explore this kind of trees and provide good solutions as fast as possible.

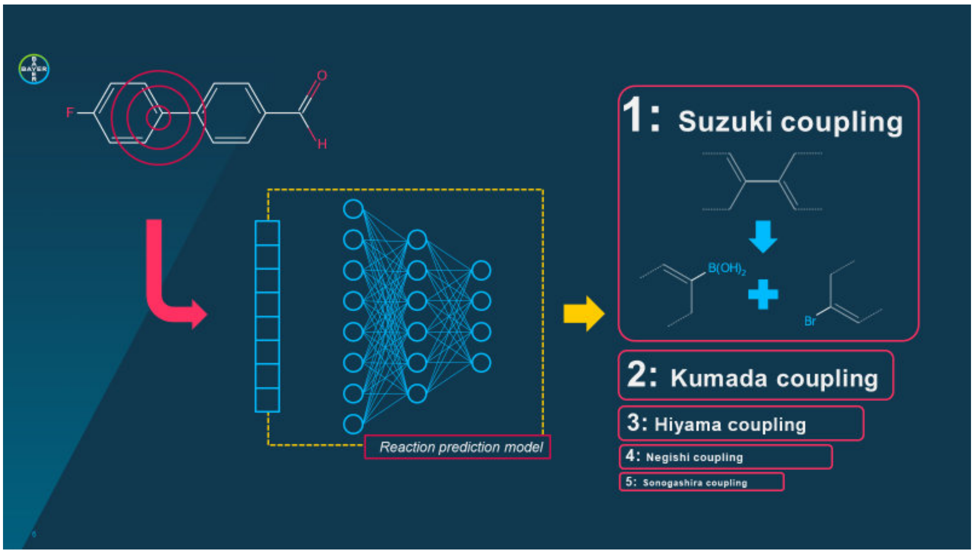


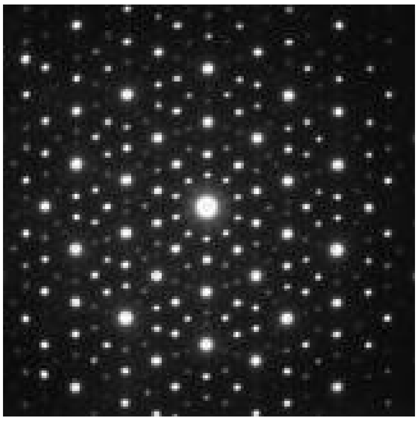
Figure 32: A neural network for the prediction of suitable reaction mechanisms

One of the nice things about mathematics is that it always surprises you where you can apply results from other sub-disciplines of mathematics. Here we use results from game theory for chemical synthesis.

The example of retro synthesis is a good way to see what the work of a modern mathematician looks like: It is interdisciplinary and team-oriented. Only through the collaboration of different disciplines can projects like this be successfully implemented. The project team includes not only mathematicians and chemists, but also chem computer scientists and computer linguists. The problems must be translated into mathematics, are then solved with mathematics, and the solution must be translated back into the technical language of the people asking the questions.

22 Quasicrystals*

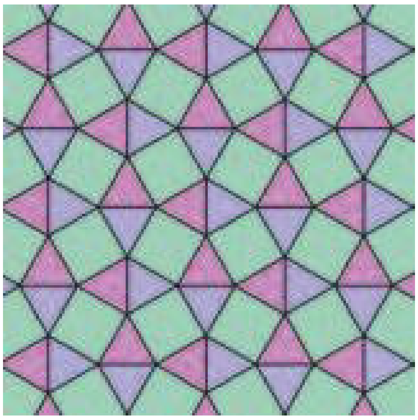
1. A Surprising Pattern



When Dan Shechtman saw this pattern under his microscope in 1982, he couldn't believe his eyes. At that time, he was studying the atomic structure of a crystal, or more precisely, a mixture of aluminum and manganese. All known crystals had a regular structure in which the pattern repeated itself over and over again. But now Shechtman saw something quite different: ten dots in ever-widening circles each time. A crystal structure that in theory could not exist at all. For all experts agreed that the atoms in crystals are in a repeating lattice, in which the distances between the atoms are always the same.

When Shechtman published his discovery and gave it the name quasicrystal, two-time Nobel laureate Linus Pauling that no quasicrystals existed, only quasi-scientists.[24]

2. Flat Fillings



For centuries, mathematicians have studied the patterns in which atomic crystals normally occur. Even though they themselves called this mosaic work: What shapes can be used to completely fill flat surfaces without leaving holes or cracks? This works, for example, with all squares or with regular hexagons. Combinations of different shapes are also possible: for example, with triangles and squares like this one. Anyway, there are all neat patterns that repeat over and over again. Just like chemists liked their crystal structures.

[24] A. Jha, *Interview: Dan Shechtman: 'Linus Pauling said I was talking nonsense'*, The Guardian, January 6, 2013, https://www.theguardian.com/science/2013/jan/06/dan-shechtman-nobel-prize-chemistry-interview.

3. A Non-Periodic Pattern

For a long time, mathematicians wondered whether it was also possible to produce a finite set of tiles that could be used to cover an infinitely large area in such a way that the pattern would never repeat. In the mid-1970s, Roger Penrose came up with an amazingly elegant answer: he works with just two different diamond-shaped tiles. In this way, you can create an infinite pattern that never repeats.

4. The Great Connection

A lattice of atoms at the corners of a Penrose control provides, under a microscope, exactly the pattern that Dan Shechtman saw. It took until 1984 for all the pieces of the puzzle to fall into place and for people to see that Shechtman was right: quasicrystals do exist. Eventually, the International Union of Crystallography even changed the official definition of a crystal: a repeating pattern was no longer necessary. In 2011, Dan Shechtman was awarded the Nobel Prize in Chemistry for his discovery of quasicrystals[25]. Steels with these irregular crystals now also have various applications. For example, they are quite wear-resistant and are used in razor blades and surgical instruments.

[25] P. Lannin, *Ridiculed crystal work wins Nobel for Israeli*, Reuters, October 5, 2011, https://www.reuters.com/article/nobel-chemistry-idUSL5E7L51U620111005

23 Best-Price Ticketing Systems: Ride First, Pay Later!**

Traveling by public transportation is made easy with modern ticketing systems: check in via app to get a travel authorization, and then just start driving. The guaranteed lowest price for the entirety of all trips for the day, even across fare boundaries and modes of transportation, is calculated after the fact.

Dr. Petra Bauer works at Siemens Technology, the central research and development department of Siemens AG, in a team of experts for mathematical optimization. She reports here on Siemens Next Generation Ticketing, a modern ticketing solution from Siemens Mobility, for which she developed mathematical algorithms.

„Which ticket do I need to travel from Zurich Stadelhofen to Winterthur? Is there a round-trip ticket that is cheaper than two individual trips? I may be traveling to a colleague's house in Oerlikon in the evening. In this case, is a day ticket worth it? How likely is it that I will actually visit her?"

Who is not familiar with these or similar considerations before starting a journey by public transport, especially in cities or regions whose fare system we are not familiar with? With modern ticketing systems, using public transportation is easy: check in for each trip at the starting stop via a ticketing app, get on, and, possibly after changing trains, get off at the destination stop. Mathematical processes take over the identification of the completed journeys and the calculation of the cheapest total price.

Figure 33: Just check in via ticketing app and go.

Check-In/Be-Out Systems

The variant of a ticketing system described above is what is known as a check-in/be-out (CiBo) system, in which the user informs the system that the trip is about to begin (check-in), but the system detects that the trip has ended (be-out).

Collected Data

In order to identify rides and calculate the best price, data is collected by the ticketing app during a ride, starting with check-in, securely transmitted to the background system, and then analyzed.

The most important of the collected data are GPS coordinates, which provide information about when the passenger was at which location. If the GPS coordinates are accurate, all means of transport are on time, and they have sufficient temporal and spatial distance, the GPS data already provide quite good information.

But what if subways are part of the transit system and GPS reception is limited there, or if we need to decide whether a slow-moving passenger is walking part of the way or sitting in a vehicle stuck in traffic?

Two other data sources help answer these and similar questions: underground, Bluetooth Low Energy (BLE) beacons installed at stations provide information, small radio transmitters that send out information that can be received by nearby smartphones. If you don't mind the cost, BLE beacons can also be installed in vehicles, for example, to better decide whether a passenger has been inside or outside a vehicle or whether he or she has changed vehicles. In addition, smartphone sensors provide data that allow conclusions to be drawn about movement patterns. One thus obtains probabilities for individual modes of locomotion („activities"), such as walking or cycling.

trip reconstruction

GPS coordinates, sightings of BLE beacons, and Activity data are referred to as „field data". Using infrastructure data (location of stops) and timetables, combinatorial optimization and stochastic methods compute the most likely sequence of timetable trips according to the data (e.g., S8, Zurich Hbf to Oerlikon, scheduled departure 9:00, Id 000011:18830:101) and other modes of travel (such as walking while transferring).

The particular challenge lies in processing very large amounts of data and dealing with uncertainties. Fog, a user who checks in a little too late, stops moved at short notice, e.g., due to an accident, are sources of error to which the processes must react robustly. The passenger knows the one, correct solution and this must be found with very high reliability.

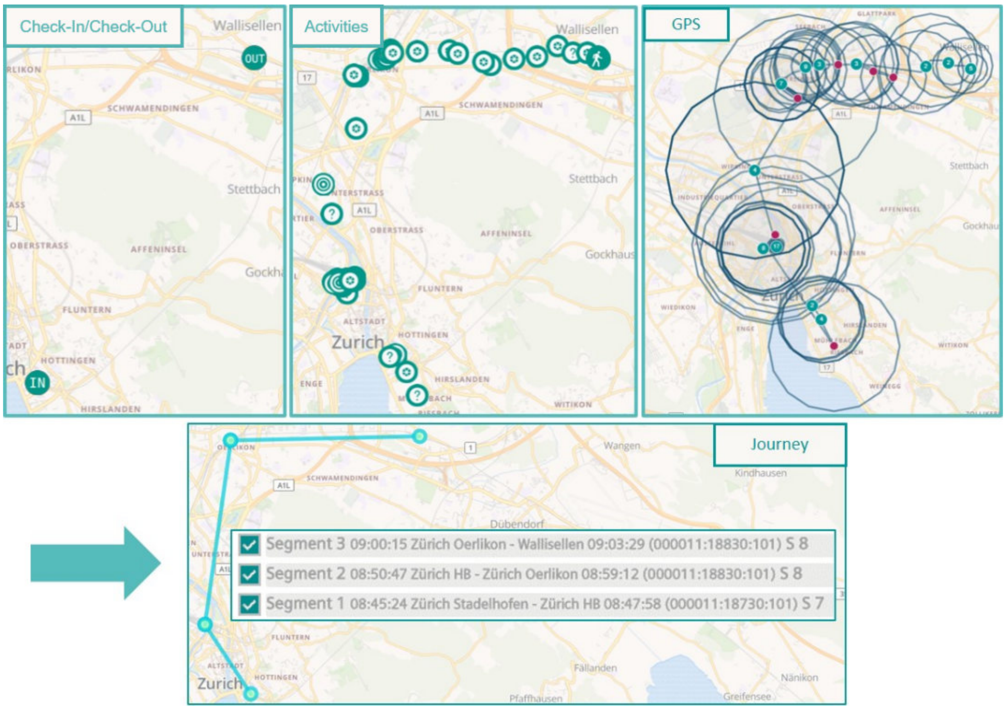


Figure 34: Top: Field data (check-in/be-out, activities, GPS coordinates); bottom: Result of the journey reconstruction.

Determining the cheapest price

The appeal of these modern ticketing systems lies in their simplicity for the user („just get going“) and the promise of being charged the lowest price after the fact.

A lot of trips can be covered with different combinations of tickets (single trips, day tickets, longer valid zone tickets, short distance tickets, special offers on certain sections, etc.). The number of possibilities quickly becomes very large, but with methods of combinatorial optimization, thanks to mathematical theory, a provably optimal solution, i.e., the most favorably priced combination of tickets, can quickly be found.

Real implementations of BiBo, CiCo and BiBo systems

In addition to CiBo systems, there are also CiCo systems (check-in/check-out) and even BiBo systems (be-in/be-out) are possible. As a general rule, the less the passenger has to do, the more equipment (e.g., beacons) is needed to achieve the required reliability of the system.

In April 2018, two systems developed by Siemens went into operation simultaneously: The world’s first BiBo system on the lines of the Südostbahn Schweiz and a Switzerland-wide CiCo system (incl. ski lifts!). The implementations serving as „proof of concept“

have proven their suitability and reliability in real operation. The further development of the CiCo system by the Siemens companies HaCon and eos.uptrade started operation in the CiBo variant in October 2020 in Osnabrück (VOS) and is available to every passenger.

It is expected that many more such systems, developed by different companies, will begin operation in the coming years.

Exkurs: The Combinatorial Optimization

Both the driving reconstruction and the determination of the best price are calculated by methods of combinatorial optimization. These procedures allow the optimal solution to be found often among trillions of possible solutions to a combinatorial problem.

A famous example of this is the Traveling Salesman Problem. A traveling salesman must find the shortest round trip through a given set of cities. Already with 65 cities he has more possibilities (64!/2) than there are estimated atoms in the universe. So to enumerate all possibilities is impossible.

Complexity theory divides combinatorial optimization problems into different complexity classes. For example, for the class of polynomially solvable problems, we know that there are fast algorithms that solve these problems optimally. For the so-called NP-hard problems (which include the Traveling Salesman Problem), one assumes that there are no methods that solve these problems quickly in general. However, it turns out that even for this complexity class, one can often compute optimal or good solutions (giving a quality guarantee) for concrete problems using mathematical methods in practice.

Combinatorial optimization has many real-world applications in logistics and transportation, manufacturing, communications, energy supply, and other areas.

24 Project Ride-Hailing Wuppertal***

As part of the joint project bergisch.smart.mobility, we are researching intelligent route planning for ride-hailing services in Wuppertal in collaboration with the Wuppertal public utility company. We use methods of multi-objective discrete optimization to determine cost-efficient and customer-friendly routes.

PD Dr. Michael Stiglmayr and Daniela Gaul M.Sc. describe their experiences with the ride-hailing project.

Ride-hailing services are seen as an important building block in the ecological urban transport of the future and represent a flexible supplement to traditional local public transport. Ride-hailing services are distinguished from line-based public transport by their adaptive route planning. Ride-hailing services differ from classic cab services in that they do not operate individual tours, but rather transport requests are bundled if necessary and several passengers share a vehicle on partial routes. In this case, one speaks of *pooling*. Due to its specific requirements, route planning for ride-hailing services requires adapted models and algorithms. In particular, time dynamics pose a major challenge to optimization algorithms: Ride requests are continuously received (online problem) and have to be answered ad-hoc. Possible decisions are the (optimized) integration of the trip into an existing tour (which can thus also be changed dynamically), the opening of a new tour, but also the rejection of the request. The primary goal of route planning is the time-efficient fulfillment of as many trip requests as possible. Unavoidable waiting times for individual passengers must be predicted as precisely as possible, in order to ensure appropriate acceptance. Economic aspects have to be considered from the beginning, since the calculated fare depends only on the starting point and the destination, while the transportation costs depend on the route taken, in particular on how well this request can be integrated into an existing tour. The trade-off between individual (user) interests and economic as well as environmental criteria is investigated with the help of multi-criteria approaches.

Instead of determining the routing directly on the road network, we use a reformulation of the problem in an event-based graph (see Figure 35). This allows us to decouple the order and distribution of passengers among vehicles from the concrete routing between two event nodes. Nodes in the event-based graph correspond to changes in the passenger occupancy of a vehicle. Since these nodes are firmly linked to an embarkation or disembarkation location and a corresponding time window, the pairwise travel times between them can be precalculated.

The ride-hailing problem can be formulated as follows: We have a total of n ride requests and m vehicles. Let $R := \{1, \dots, n\}$ be the set of ride requests. Each request i from R contains an entry location i^+ and an exit location i^-. Each request also includes the number of people to be transported, the maximum travel time, and an boarding and an alighting time window.

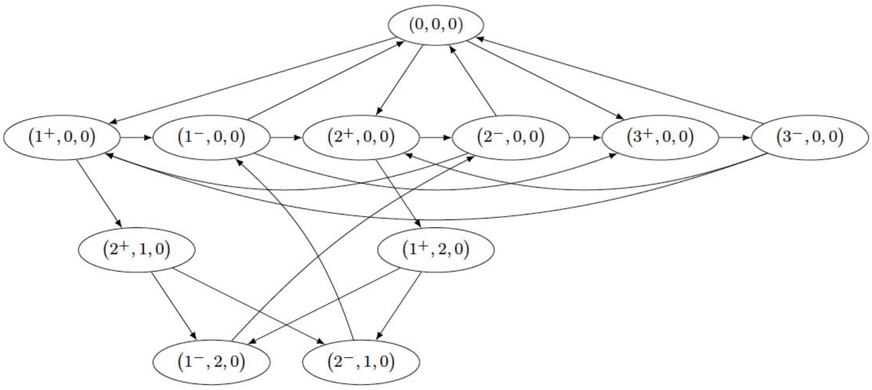


Figure 35: Event-based graph for vehicles with three seats.

In the following example, we assume 3 seats in a vehicle for simplicity. The event-based graph consists of nodes and edges. We write a node as a 3-tuple $v = (v_1, v_2, v_3)$ where each entry describes one seat of the vehicle. If $v_j = 0$, then the seat i is empty. The node $v = (1^+, 2, 0)$ describes a vehicle with customers 1 and 2. The first coordinate of the node additionally indicates which location was visited last (in this example customer 1 got on). A node is only added to the graph if the sum of requested seats does not exceed the vehicle capacity. In Figure 35, the event-based graph is shown for a scenario with three customers, where customers 1 and 2 both request two seats.

Using the event-based graph, the ride-hailing problem can be described as a mixed integer linear optimization problem (MILP). That is, we describe the set of admissible solutions to the ride-hailing problem by linear inequalities with integer and continuous real variables. Here, the integer variables x_α model the discrete decisions in which order which vehicle picks up or drops off which passengers. The continuous variables B_v contain the times at which a vehicle arrives at the nodes (boarding/disembarking). These must be within corresponding time windows. In addition to the constraints that ensure that passengers board and alight within the specified time windows, a number of other constraints are required that model, among other things, the distribution of passengers among vehicles and the order sequence of vehicles. The objective function evaluates the quality of an admissible solution. In the simplified model in Figure 36, the total distance traveled by all vehicles is used for this purpose. Thus, an optimal solution for this model would be a route scheduling that satisfies all constraints while having the lowest total travel distance.

If one does not want to limit oneself to purely economic aspects for the evaluation of a solution, further criteria can be considered in addition to the minimization of travel costs, which the objective function in Figure 36 describes. These can be customer-

$$
\begin{aligned}
\min \ & \sum_{a \in A} c_a \, x_a \\
s.\,t. \ & \sum_{a \in \delta^{in}(v)} x_a - \sum_{a \in \delta^{out}(v)} x_a = 0 \quad \forall v \in V, \\
& \sum_{\substack{a \in \delta^{in}(v) \\ v \in V_{i+}}} x_a = 1 \quad \forall i \in R, \\
& \sum_{a \in \delta^{out}(\mathbf{0})} x_a \leq |K|, \\
& e_0 \leq B_{\mathbf{0}} \leq \ell_0, \\
& e_{i+} + (\ell_{i+} - e_{i+}) \left(1 - \sum_{a \in \delta^{in}(v)} x_a \right) \leq B_v \leq \ell_{i+} \quad \forall i \in R, \, v \in V_{i+}, \\
& e_{i-} \leq B_v \leq e_{i+} + L_i + s_{i+} + (\ell_{i+} - e_{i+}) \sum_{a \in \delta^{in}(v)} x_a \quad \forall i \in R, \, v \in V_{i-}, \\
& B_w - B_v - s_{i+} \leq L_i \quad \forall i \in R, \, v \in V_{i+}, \, w \in V_{i-}, \\
& B_w \geq B_v + s_{v_1} + t_{(v,w)} - \tilde{M}_{v,w} \, (1 - x(v,w)) \quad \forall (v,w) \in A, \\
& x_a \in \{0,1\} \quad \forall a \in A, \\
& B_v \geq 0 \quad \forall v \in V.
\end{aligned}
$$

Figure 36: Simplified ride-hailing optimization problem.

oriented criteria, such as minimizing the additional travel time compared to a direct trip, or the waiting time at the boarding point. Furthermore, situations may arise where some requests have to be rejected if serving these requests means large additional detours for the remaining customers. Another criterion is therefore the fulfillment of as many requests as possible. When considering several criteria (objective functions), one speaks of multi-criteria optimization. When solving a multi-criteria optimization problem, one wants to find so-called Pareto-optimal solutions. These are solutions where one cannot improve in any criterion without getting worse in another criterion.

For exact solutions to the mixed-integer optimization problems that arise, we use branch-and-bound algorithms. These use recursive partitioning of the decision space (set of admissible solutions) with respect to integer variables to detect and reject suboptimal decisions early using bounds. For a small example with six customers and two vehicles, the optimal route planning is shown in Figure 37.

. For large instances, however, the exact solution may be too time consuming, which is why we are currently working on complementary heuristic methods that can quickly generate a good, but not necessarily optimal, solution. This is especially relevant in the dynamic view of the problem, when driving requests arrive successively, have to be ans-

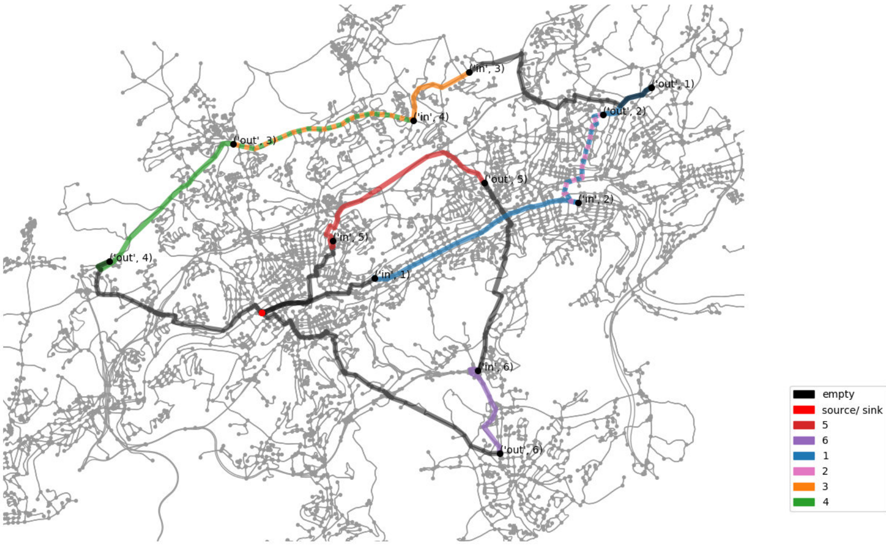


Figure 37: Route planning for 6 customers and 2 vehicles (on the dashed sections several customers share one vehicle).

wered and integrated into the route planning. For this purpose, one can use the static models discussed above and integrate them into a so-called *rolling horizon approach*. In addition, however, static model solutions combined with simulated trip demand data also provide important information for the business planning of ride-hailing services. For example, questions about potential area expansions or comparative studies on the replacement of bus routes with ride-hailing services can be investigated.

25 The E-Bike and its Drive*

There is a lot of mathematics from many subsections in an e-bike.

Dr. Markus Hinterkausen, Dr. Uwe Vollmer and Dr. Uwe Iben from the Corporate Sector Research and Advance Engineering of Robert Bosch GmbH summarize their experience with e-bike development:

„Powerful, quietly humming and innovative in design – eBikes are real trend products of the 21st century. However, they've been around much longer than we'd like to believe. A first patent for an electric bicycle was filed as early as 1895. E-mobility was already seen as the future mobility in cities around 1835. Smooth roads and short distances were the arguments. As early as 1900, more electric bicycles were patented, built and marketed. This was at the same time as bicycles with steam engines, which, however, did not catch on due to their heavy weight. But the e-bike also remained a marginal phenomenon until 2010. The high weight of the batteries, the short range and the jerky drive were the reasons. But with the lithium-ion battery and new small sensors, a new era of e-mobility began. In 2009, Bosch eBike Systems entered e-mobility as a start-up within the Bosch Group. The success story began with the first eBike drive system, which was launched in 2011 and makes relaxed cruising in the city just as possible as sporty biking in nature. A major milestone was the market launch of ABS for e-bikes in 2017 [26].“

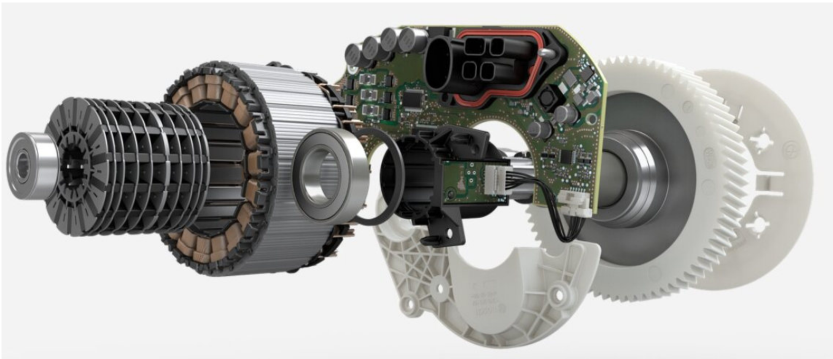


Figure 38: Electric motor, sensors and transmission (source: Bosch eBike Systems).

„A success story has many reasons. One of them is the targeted development of the electric motor, transmission and sensor technology components. They form the heart of the drive, see Figure 38. Another reason is the user-friendliness of the system. Here, for example, a special on-board computer was developed that can be connected to a smartphone.“

[26] https://www.bosch-ebike.com/en/everything-about-the-ebike/stories/200-years-of-bicycle-history

„A variety of mathematical methods are used in the development of the e-bike drive and its components. Analytical methods are used for the initial design of the drive and the e-machine. With these approaches, different designs with their characteristics are obtained very quickly. Mathematical methods used here are vector and matrix calculations from linear algebra and, at least at one point, conformal mapping. These are angle-true mappings, which are used in physics, electrical engineering or fluid mechanics, among others. This is where initial customer requirements are incorporated, such as torque characteristics, etc. Methods of stochastics help in the tolerance analysis of components, in order to include manufacturing processes and their tolerances in the design of the powertrain right from the start. "

„In a further step, finite element methods (FEM) and boundary element methods (BEM) are applied to refine and analyze the drive design. Here, the resulting complex systems of equations are solved numerically. The weighted residuals approach with Galerkin test functions is usually used. In a first step, the magnetic circuit is modeled and simulated using these numerical methods in order to observe the operating behavior at different operating points. From these calculations, behavior models in the time domain or in the frequency domain are derived. Laplace and Fourier transforms are used for this purpose. Models in the frequency domain are especially useful for evaluating acoustic behavior, since an important customer requirement is low pleasant noise. "

„The behavior models in the time domain are used, for example, to design the control of the drive and to simulate driving cycles. In the time domain, nonlinear effects can be considered, e.g. in material behavior. Gradient-based methods for topology optimization are used in the course of drive design to improve e. g. thermal, mechanical, electrodynamic or acoustic properties. They usually require the use of powerful computers. Multi-objective optimization by searching for Pareto-optimal designs in several physical domains such as structural strength, electrodynamics, and also manufacturing and material costs are used to find the best designs for the drive. The controls are designed in the frequency domain using system description by Laplace transforms and in the time domain by differential equations."

„To this end, tools such as the Bode diagram or the root locus curve are used to evaluate the quality and stability of the control loop. The method of transfer path analysis describes the path of structure-borne sound from the source to the enclosure, from where it reaches our ear as airborne sound. Modern methods from the field of artificial intelligence (AI), such as deep learning, support the evaluation of sound with respect to psychoacoustic perception."

„All these mathematical methods are applied to the design of the electric drive for the e-bike. Step by step, the methods and procedures are improved and extended. In addition to accuracy, the focus is on the predictability and computational speed of the algorithms. At the same time, efforts are being made to incorporate the findings from earlier designs into new ones. Piece by piece, the methods and algorithms are being combined to form a digital twin, which in the future will map the powertrain completely digitally, opening up completely new possibilities for the engineer to evaluate functions and costs. "

26 The Surprise Delay**

An additional road can lead to a longer travel time. This counterintuitive result is called the Braess paradox and occurs in all sorts of forms.

In the *Networks* project, a team of scientists is researching the transport networks of the future. Here, the *Braess paradox*, from the stochastic networks plays a major role. Stochastic networks are networks in which randomness plays a role. In these networks, the traffic flow, just like in road traffic, can get worse if you increase the capacity of the network. This seems to be a paradox, but once you understand the mathematics, you see that there really is no contradiction.

The Old Situation

A small example makes clear what happens. Ten vehicles want to drive from source $\mathbf{Q}$ to destination $\mathbf{Z}$. Each driver can choose between two routes: the lower route via $\mathbf{A}$, or the upper route via $\mathbf{B}$. The travel time for each route depends on the number of vehicles V that pass, see the drawing for the exact times. For example, the trip from source to $\mathbf{A}$ costs twice the number of cars choosing this route (in minutes). If all 10 cars choose this route, then the travel time on this piece will be twenty minutes for all of them. If only one car travels, it will reach $\mathbf{A}$ in two minutes.

In an equilibrium situation, the travel times over the lower and upper legs are exactly the same and no driver has any interest in changing roads. In this case, five cars take the upper route and five take the lower route. This solution is also obvious because of the symmetry of this problem. It is easy to calculate that in this case everyone has a driving time of 30 minutes.

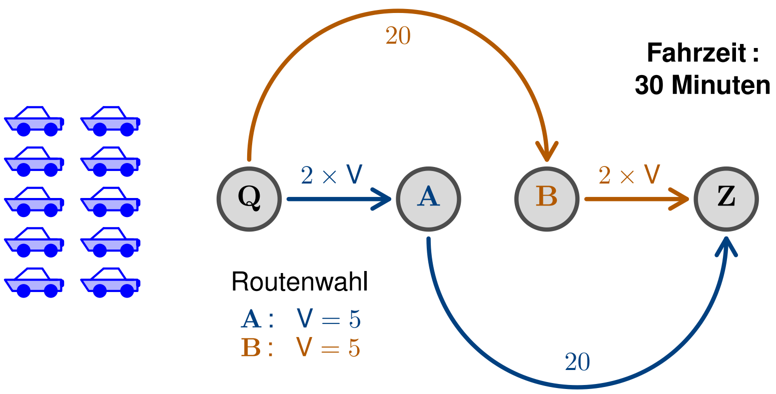


Figure 39: Alte Fahrsituation

"Improvements"

Then the government decides to build a new fast road from $\mathbf{A}$ to $\mathbf{B}$ (trivial). At source $\mathbf{Q}$, the ten drivers can now choose from three routes: The two old ones plus the new road that runs over $\mathbf{A}$ and $\mathbf{B}$. Now, if one of the 5 cars on the lower route takes the new road, its travel time drops to 22 minutes. This slightly increases the travel time for the 5 cars on the upper route: it is 32 minutes.

Therefore, it is wiser if one of them also takes the new path via $\mathbf{A}$ and $\mathbf{B}$. After that, there is another equilibrium situation where the three routes each take the same amount of time. The travel time is now 40 minutes for all of them, which is ten minutes longer than before the additional road was added. At this point, however, no driver can reduce their travel time by individually taking a different route.

This is a miraculous phenomenon. It would be better if everyone ignored the new road, but then there will always be a driver who thinks he will travel faster if he takes it. For a while he seems to be doing better, but in the end everyone is the victim. This effect is inevitable. It would be much more efficient if people could decide centrally which car goes where.

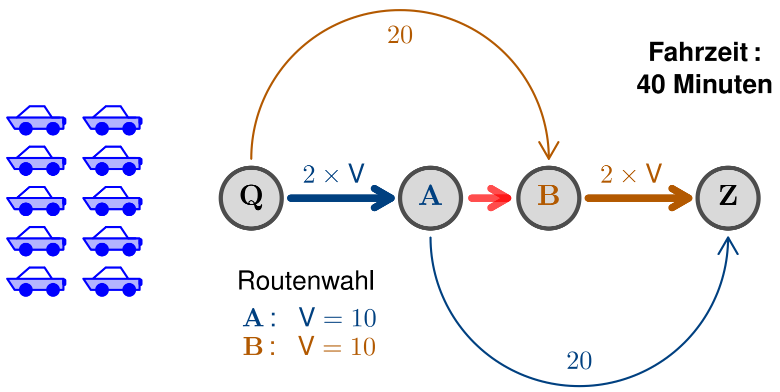


Figure 40: Driving situation with the additional road.

The Braess Paradox is a regular phenomenon. In Boston, commuters got a longer commute when an additional tunnel was built under the name Big Dig. The opposite is also true: when New York temporarily closed the 42nd Street arterial, traffic seemed to flow more smoothly through the city. The phenomenon has also been observed and documented in Stuttgart or Seoul.

Future Plans

Researchers plan to better map such behavior in transportation networks in the coming years and also look at how to improve the system by tracking a small percentage of cars. They will also work on networks for communications, energy and logistics. The beauty of this is that all the networks are broadly the same structure. So the problems to be addressed are not reserved for just one application area.

Figure 41: Traffic jam at Robert-Daum-Platz in Wuppertal. In the background, the University of Wuppertal. Photo: Tomas Riehle (arturimages)

27 Modeling, Simulation and Optimization for Noise Reduction in Disc Brakes***

Model-based design and improvement of advanced vehicles. A success story of cooperation between mathematics and engineering.

Modern industrial products such as automobiles, trains or airplanes are highly complex systems based on a combination of different fields of engineering sciences, e.g. structural mechanics, fluid dynamics, electrical engineering, hydraulics, process engineering. In the design and optimization of such systems, the use of numerical simulation based on mathematical models and the optimization of the system behavior and the manufacturing process plays a central role today. If one wants to describe such complex systems by mathematical models, one is forced to make compromises regarding the modeling quality, between accuracy and computation time during the simulation, and the achievable optimality. Mathematics provides the tools to do this in a meaningful and rational way by analyzing models and developing mathematical algorithms, quantifying modeling and approximation errors, and estimating complexity and computation time.

This approach to a real problem is described by Prof. Dr. Volker Mehrmann (Institute of Mathematics, TU Berlin) using the example of the design and improvement of the noise emission of disc brakes. For a mathematically and mechanically detailed description, see this article[27].

Disc brake squeal in motor vehicles (but also in brakes of bicycles or trains) is an unpleasant noise. A major cause of squealing is self-excited vibration of the brake disc, which is generated by the frictional forces at the interface between the brake disc and the brake shoe. Figure 42 shows a model of an entire disc brake including the suspension.

In today's industrial practice, such a brake is analyzed and improved by laboratory experiments and by numerical simulation based on a finite element model. For this purpose, the structure is decomposed into very small cuboids or tetrahedra (finite elements) and the deflection from a state of rest, or rather the velocities of motion are approximated by piecewise polynomials on these finite elements, see Figure 43 for such a decomposition into finite elements. This results in a very large system of differential equations. For each finite element six or more degrees of freedom are used (e.g. three for the position and three for the velocity). For a real brake, this can be millions of degrees of freedom, depending on how accurate the description needs to be. A big problem is that the disk is of course rotating very fast and thus the coordinates have to be considered in a co-moving coordinate system. However, in addition to modeling the structure using finite elements, modeling the frictional behavior in particular is a very difficult problem

[27] N. Gräbner, V. Mehrmann, S. Quraishi, C. Schröder, U. von Wagner, *Numerical methods for parametric model reduction in the simulation of disc brake squeal*, Zeitschrift für Angewandte Mathematik und Mechanik Vol. 96 (2016), 1388–1405.

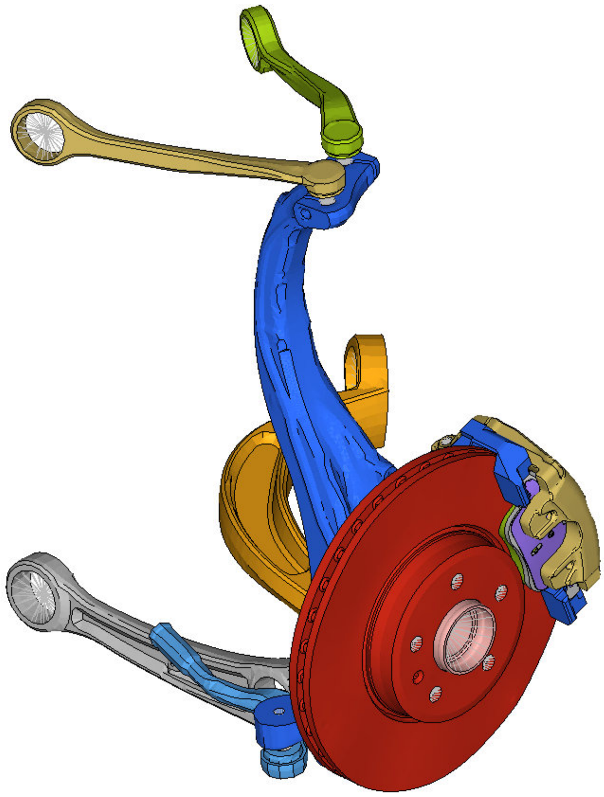


Figure 42: Brake model.

because friction occurs at the molecular level. Therefore, for the modeling of friction, one uses surrogate models that are obtained by fitting parameters to measurements from experiments.

Overall, the currently industrially used modeling results in a linear second-order differential equation of the form

$$M_\omega \ddot{u} + D_\omega \dot{u} + K_\omega u = f,$$

where M_ω, D_ω, and K_ω are very large matrices (size $\approx$ number of degrees of freedom) describing the relationship between acceleration, velocity, and position coordinates. These coefficient matrices depend on different parameters, e.g. material parameters of the brake, abrasion of the brake disk, internal damping etc.. Here only the dependence on the rotational speed ω of the brake disc is indicated, which is varied in a certain parameter range. In contrast to classical models for vibrations of beams or plates, here circulatory and gyroscopic forces have to be taken into account due to the rotation of the disk, so that the matrices D_ω and K_ω are non-symmetric, while M_ω is a symmetric mass matrix. Moreover, for self-excited oscillations, the external force f is neglected.

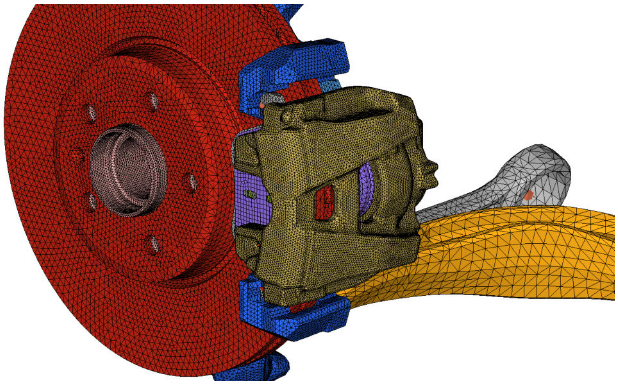


Figure 43: Example of a finite element model of a disc brake.

In this model description, of course, a large number of simplifications have been made and a comparison with laboratory experiments for the evaluation or validation of the model is necessary, where also the goodness of the model (more or less good) is quantified.

To understand the dynamics of the system, resp. analyze and change oscillation modes of the differential equation, one first makes a so-called eigenvalue analysis, i.e. one calculates on the basis of a Fourier analysis, eigenfrequencies $\lambda(\omega)$ and eigenvectors $x(\omega)$, so that for given ω the quadratic eigenvalue equation

$$\left(\lambda(\omega)^2 M_\omega + \lambda(\omega) D_\omega + K_\omega\right) x(\omega) = 0$$

holds. This eigenvalue analysis holds locally near an equilibrium state, and the system is unstable (and will squeak more or less audibly) if there are eigenvalues with positive real part for certain ω.

There are now $2n$ eigenvalues for this system (where n is the system size). Even with today's computer clusters, it is not realistic or practical to compute all of these eigenvalues and associated eigenmodes for many parameter values ω to determine at what ω the squeak starts and at what frequency it squeaks, and then optimize the model based on variations in geometry, mass distribution, or material parameters.

For the calculation of the eigenvalues modern algorithms of the numerical linear algebra (so-called Krylov space methods) are used, with which one can preset ranges in which one searches eigenvalues. That means one calculates only the eigenvalues in a certain area of the complex number plane. These methods are very well examined concerning their error susceptibility and one can estimate the errors in the approximation methods and also adapt them to the modeling errors, so that one can carry out an estimation for the total error of modeling and eigenvalue calculation errors.

The analysis of natural frequencies and oscillation modes is thus possible, but extremely computationally intensive, especially if optimization is then to be performed via the parameters. For this reason, a very small parameter-dependent reduced model is first generated, which can be used for parameter optimization. There are a variety of model reduction methods. For this purpose, the large problem is projected onto a small problem by searching the eigenvectors $x(\omega)$ in a subspace spanned by the orthonormal columns of an $n \times d$ matrix Q, where d is of the order of a few hundred at most. The classical method (modal truncation) in industrial use uses the eigenvectors of the matrix pair $\lambda M + K_1$ where K_1 is a symmetric part of K to the largest eigenvalues (these are all real and negative). In a modern approach[1], where the errors are much smaller, a matrix Q is computed such that, in a large parameter range, the relevant eigenvalues of the reduced model (i.e., those with positive real part), well approximate those of the full model, and this with an error estimate. For this purpose, the relevant eigenvalues and eigenvectors of the full model are computed for some randomly chosen ω values in the range of investigation, and then a singular value decomposition (Proper Orthogonal Decomposition (POD)) is used to generate from them a small space that can be used in the whole parameter range.

Of course, another model reduction error is made during model reduction, which should be of the same order of magnitude as the other errors made so far if the reduction method is good.

An illustration of different error measures as a function of the dimension of the reduced model for two industrial finite element models M_1 and M_2 and the classical as well as the new method is shown in Figure 44.

One can see very nicely that the POD method, with a very small d produces very small errors in the eigenvalues err_λ and eigenvectors err_x. Moreover, to these errors there is an error estimation, which allows to adaptively control the size d and the number of chosen evaluation points. In comparison, the error with the classical method stagnates even with an enlargement of the space.

In addition to the error estimation, each evaluation step in the parameter optimization is then also much cheaper (on average 30 times faster than with the classically reduced model), if one solves the reduced eigenvalue problem

$$\left(\lambda(\omega)^2 Q^\top M_\omega Q + \lambda(\omega) Q^\top D_\omega Q + Q^\top K_\omega Q\right) x_d(\omega) = 0.$$

This can now be used to efficiently improve disc brakes, and the method has also led to the discovery of instabilities (eigenvalues with positive real parts) in real examples that the classical model reduction approach had not found.

The new method was implemented as an open source Python script and is freely available in general.

This example, but also a large number of other applications, illustrates the contributions modern mathematics is making to improve methods and understanding in all areas of science.

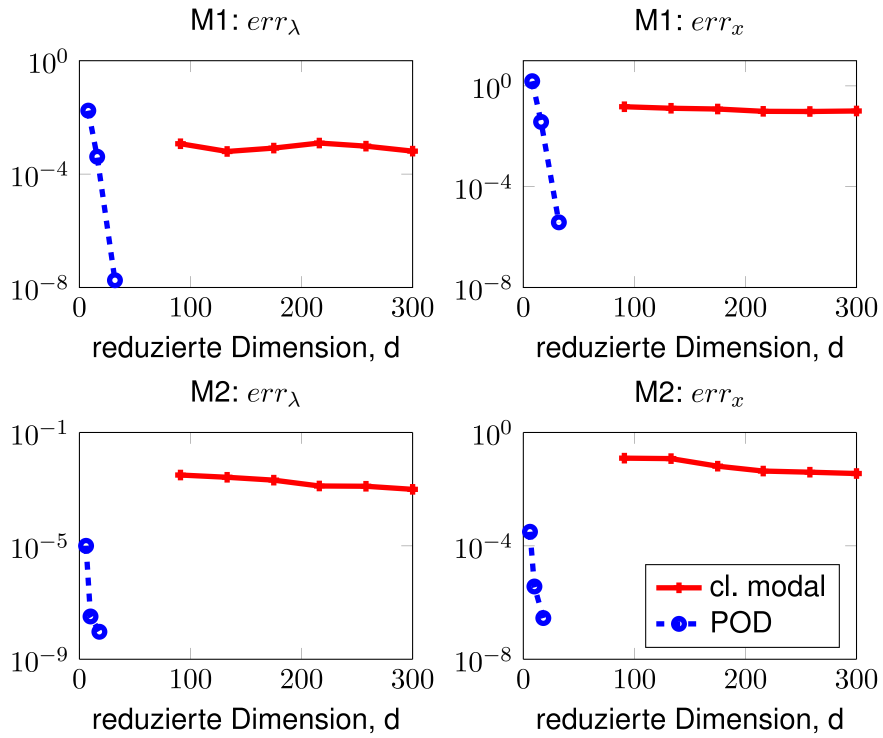


Figure 44: Dependence of error quality on reduced dimension d for two industrial models M_1 and M_2, and classical and modern model reduction methods.

28 The Ambulance will arrive in 15 Minutes**

How many ambulances are needed to ensure that no one has to wait more than 15 minutes for emergency care? A 1909 mathematical model provides a solution to this staffing problem.

For urgent and life-threatening situations, such as traffic accidents and heart attacks, an ambulance must be on the scene within fifteen minutes of calling 112. For non-life threatening situations, such as a broken leg, there is a maximum approach time of half an hour.

But that's all easier said than done. How many ambulances does a city or region need to meet these requirements? If half the ambulances have nothing to do all day, it's clearly too many. But if half the ambulances are late by default, there are clearly too few ambulances. So where is the optimum? And where can those ambulances be at their best? Certainly not all in one pile, but perhaps not all as far apart as possible.

Timeless Math

To determine the optimal number of ambulances, one needs two important parameters. These are the number of calls received per minute and the length of time an ambulance is staffed at a given time of day. Ambulance services continuously collect data on this topic. Mathematicians can use this data to optimize ambulance scheduling.

The easiest way to determine the optimal number of ambulances is to use a formula established by Danish mathematician Agner Erlang in two articles in 1909 and 1917:

the *Erlang C formula*. Erlang used probability theory to design the formula for the then-emerging fixed-line telephone services. In his first paper in 1909[28] he proved that randomly distributed telephone calls follow *Poisson's distribution law*. Depending on the number of expected telephone calls per minute and the average call duration, he determined the number of telephone lines needed such that the probability of someone not calling is below a certain quality value. His second work[29] contained formulas for loss and waiting time („Erlang-B" and „Erlang-C"), which are well known today in the theory of telephone traffic.

The beauty is that the same 1909/1917 formula can be applied in many different areas today: From determining the number of towers needed in a cellular network, to determining the network capacity for streaming video, to determining the optimal number of ambulances, or the number of employees in a call center to guarantee a maximum-acceptable customer waiting time (a so-called service level). This wonderfully demonstrates the timelessness of mathematics.

Shadow Planning

Researchers are working closely with people on the ground: emergency services, fire departments and police forces. They built software that allowed them to keep a shadow plan for a year. In this shadow plan, they saw their own manual planning, our mathematically calculated planning, and real-world ambulance deployment. After a year, they concluded that their planning was more accurate than that of the ambulance services. This means that they can provide more quality with the same number of ambulances and people, or provide the same quality at a lower cost.

Scientists are making their mathematical models more and more realistic and sophisticated. One new branch of this is dynamic ambulance management. Typically, ambulances return to a fixed location after an urgent assignment. In general, it is better for the service industry not to return ambulances to a fixed location, but to determine which is the most convenient temporary waiting location at that time. This depends on where the other ambulances are at that time and the expected number of disasters at a particular location.

However, what is mathematically optimal may not always be feasible in practice. Mathematically, all ambulances would have to be moved a little to the operations center after each emergency call, but of course that is also undesirable. By consulting with people in daily practice, the models are adjusted as closely as possible to their needs, such as break times, or the heterogeneity of the staff.

[28] A.K. Erlang, *The Theory of Probabilities and Telephone Conversations*, Nyt Tidsskrift for Matematik B, vol 20, 1909.

[29] A.K. Erlang, *Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges*, Elektrotkeknikeren, vol 13, 1917.

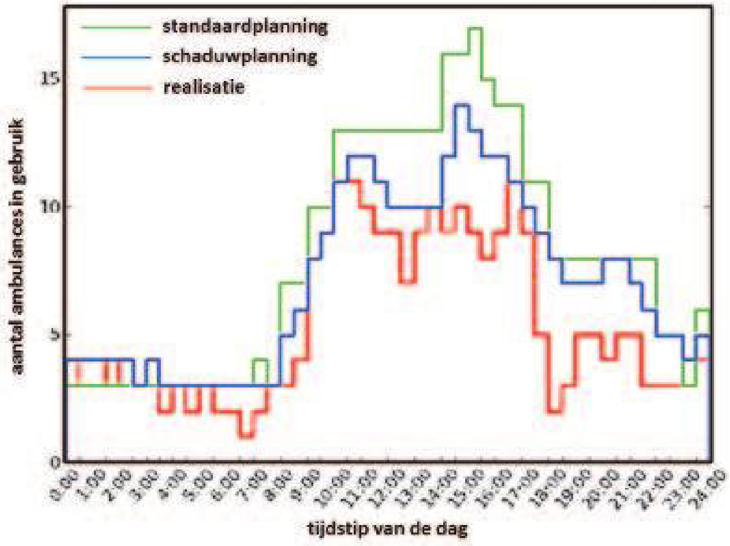


Figure 45: The mathematically calculated shadow planning works better than the planning used in practice.

29 Optimal Boarding at the Airport**

Markov-Chain Monte-Carlo methods can be used to simulate the boarding of an airplane and thus optimize it and save time.

Who doesn't know the problem: When boarding the plane, a traffic jam quickly develops and the so-called boarding takes (too) long. This not only annoys passengers – boarding time is also an economic factor for airlines. Therefore, airlines try to minimize boarding time with a wide variety of boarding strategies. Often, passengers sitting in the back are supposed to board first. This is the so-called back-to-front strategy. Similarly, it seems reasonable to 'board' people with a window seat first, then a middle seat, and finally those with an aisle seat, the so-called outside-in strategy. One could also divide passengers into different classes (e.g., by booking class, FrequentFlyer or age, etc. (so-called multiple classes strategy) or not use any strategy at all (so-called random boarding). Of course, a combination of these strategies is also possible.

It's the usual tale of woe for many travelers waiting to board a plane. In the process, the airline has to go through a list of passengers who have priority: People with disabilities (who rely on a wheelchair, for example), families with children, first and business class passengers, frequent flyer card holders and passengers who have paid for priority seating. When economy class travelers are called, they stand in line in the cabin and have to wait until those in front of them have stowed their luggage and sat down. By then, the overhead bins are often full.

To generate more revenue, airlines have introduced new fees for priority boarding and checked baggage. The latter has only exacerbated the problem because passengers are now bringing more wheeled suitcases on board, blocking aisles during boarding. According to research by Boeing, boarding time has increased significantly since the 1970s[30] and, as a result, airlines are interested in reducing boarding time, preferably without loss of revenue and inconvenience to passengers.

The first idea was to charge a slightly higher fee for hand baggage than for checked baggage, and resulted in a shorter boarding time, but with dissatisfaction among customers who could not understand this pricing policy. Therefore, other boarding strategies were developed to leave the gate on time. For example, passengers in the back rows may be allowed to board first, or those with window seats may be given priority.

Another ingenious strategy has come to mind, including one no longer in use known as „inverted pyramid“ (engl. Reverse Pyramid (RP)). The reverse pyramid boarding method was developed by Van den Briel et al.[31] developed. This method separates passengers into boarding groups based on the position of their aircraft seats. The boarding groups are formed according to a „diagonal load“ scheme. This means that in a group, most

[30] https://www.boeing.com/commercial/aeromagazine/aero_01/textonly/t01txt.html

[31] M.H.L. van den Briel, J.R. Villalobos G.L. Hogg GL, T. Lindemann A.V. Mulé, *America West Airlines Develops Efficient Boarding Strategies*, Interfaces 35 (2005), 191-201.

passengers have seats in the back of the aircraft, some in the middle section, and a few in the front, with passengers in the middle/front section sitting closer to the windows. Within each boarding group (3 to 6 groups), passengers board the aircraft randomly. This boarding method has since been discontinued (because business passengers found no room for their luggage at the end of the pyramid), but is experiencing a revival in the era of the Corona pandemic and associated distance rules, as this simple RP strategy can be used to distribute passengers more evenly throughout the cabin, and allows passengers to find their seats more quickly[32]. This RP method also helps passengers stow their luggage closer to their seats, so fewer bags have to be checked at the last minute – a common cause of delayed flights.

By so-called „Markov-Chain-Monte-Carlo optimization methods“, a mathematical algorithm for random events, one can simulate the boarding process and also optimize it. For example, the exact seating plan of the aircraft matters, and whether it is a narrow-body aircraft like the Boeing 737 with only one aisle, or a larger aircraft for transatlantic flights where passengers have two possible paths. Better than standard boarding in the simulation is random boarding, where passengers board purely at random. Another approach is to divide passengers into boarding groups based on when they check in. This reduces delays caused by seat assignments. However, most passengers have already reserved a seat online before they get to the airport.

[32] R.J. Milne, et al., *Adapting the reverse pyramid airplane boarding method for social distancing in times of COVID-19*, PloS ONE 15(11) (2020), e0242131.

30 Mathematics in Modeling for Systems Analysis*

How should the energy system evolve? What are the connections between our consumption and emissions? Mathematical models can help answer such questions in systems analysis.

Modeling always helps us when relationships are too complex to understand intuitively. In this context, we speak of system analysis. For example, hydrogen is an important component of the energy supply of the future. But green hydrogen can only be produced if sufficient green electricity is available. What then is a sensible strategy, should we build electrolysers for hydrogen production as soon as possible? What interactions will then result, will electricity then become more expensive in Germany? Or will perhaps even more emissions be caused if more electricity is generated from coal or purchased from abroad as a result? To answer such questions, we can use models based on mathematical methods to represent and investigate the interrelationships.

Here we show two examples of how different mathematical models are used in system analysis: Optimizing Energy System Models to determine cost-optimal developments and Input-Output Modeling to map macroeconomic impacts of consumption and production.

Figure 46: The interactions and relationships in the energy system are complex and not always intuitive to understand – this is where models can help.

Optimizing Energy System Models

The goal of such models is to answer questions such as „What should be the most cost-effective way to meet electricity demand over the next few decades?“, „How might the power supply structure change if one assumes that carbon allowance prices rise more sharply?“, or „How is electricity storage optimally used?“. For this, researchers like Christine Krüger and José Acosta Fernandez from the Wuppertal Institute for Climate, Environment, Energy are building a model in which the components of the energy system are represented by equations.

Electricity generation from wind energy, for example, is a time series that is derived from the installed capacity of the wind turbines and the course of the wind speed. A cogeneration plant is a unit that converts a fuel such as coal and converts it into electricity, heat, and emissions, a power line is described by the power it can transport and the losses it incurs. In this way, all the relationships of the system are put into equations. Boundary conditions are also formulated as equations: Electricity demand must always be met at any given hour, and emissions must fall within a given range. An important element of such a model are the costs: fuel prices, prices for emission certificates, costs for maintenance and so on.

This system of equations is then used to formulate an optimization problem: Minimize the cost of the overall system, and do so in a way that satisfies the constraints. The result is then the most favorable system under the assumptions made. And one can then play with these assumptions: How high does the emissions price have to be so that electricity generation would be climate neutral in 2040? How does the expansion and use of power plants change if one assumes a different development of fuel prices? Or what are the effects of specifying minimum or maximum quantities for hydrogen? Energy suppliers, for example, can then derive their strategies from such studies, or recommendations can be formulated for policymakers.

Advanced Input-Output Models

Models are also used to quantify the various impacts of human production and consumption activities, whether desired and undesired. These models are used to examine both the causes and ways to change the magnitude of these impacts. It will also be used to define appropriate indicators to track changes over time and then to design appropriate policies based on these changes.

For greenhouse gases, one such indicator is, for example, the „carbon footprint (CF)“. To calculate this, the „Extended Input-Output Analysis (E-IOA)“ is applied. From a consumption perspective, the carbon footprint expresses the total greenhouse gas emissions caused by the use (consumption) of a product group along the global supply and manufacturing chain. Thus, all greenhouse gases are included that are produced directly, but also indirectly, along the entire production chain. The interdependence of these production activities can span all sectors of the economy. These interdependencies, i.e., the mutual interrelationships, of the various economic sectors are at the heart of

E-IOA. An example of this is: car use → manufacturing of cars → manufacturing of various machines → production of steel for the cars and the machines → production of plastics → mining of minerals, coal, petroleum, gasoline → manufacturing of trucks to transport the minerals ...

The basis of E-IOA is „Extended Input-Output Tables (E-IOT)“. An E-IOT is basically a matrix in whose rows and columns the interrelationships of services and products in an economy are mapped in detail and given numbers – for example, what is the increase in energy consumption if steel production is increased by 5%? In addition, there are socio-economic and environmental direct effects (e.g. greenhouse gases). The complex interrelationships lead to very large matrices, which require suitable mathematical methods to handle.

These are two examples of the large world of mathematical models in energy system analysis. One thing is true for all models: a model is not reality; it is always a simplified, abstract representation. The art lies in designing a model in such a way that, despite this limitation, it can provide an answer that is appropriate to the question at hand. That is why there is not only a multitude of models for the most diverse purposes, but also usually several models for the same object.

31 Climate Change: the Optimal CO_2 Tax*

Climate change is a pervasive, much debated, and global problem. Research regarding climate, its evolution, and its governance is being advanced in all walks of life. One current policy measure to combat climate change is the implementation of a CO_2 tax. Mathematics plays a central role in the implementation of a CO_2 tax.

Leon Hoffe and Benjamin Leonhardt are master's students in business mathematics at the University of Wuppertal. As part of their bachelor theses, they dealt with analytical and numerical methods for determining the optimal pricing of a CO_2 tax.

Responsible for the climate on our earth is the greenhouse effect. The greenhouse effect is triggered by greenhouse gases. Among other things, these are carbon dioxide (CO_2), methane (CH_4) and nitrous oxide (N_2O). Since the industrial revolution, the concentration of greenhouse gases in the air has been steadily increasing. Accordingly, the global mean temperature has also been rising steadily since industrialization. The extent of climate change is now so great that it is perceived by diverse sectors of the world's population. This is recognized by political agreements such as the Paris Climate Agreement, which was adopted in 2015. In addition, the scale is becoming more present in news and social media or through worldwide Fridays for Future demonstrations. The predominantly youthful participants of the demonstrations advocate for fast, comprehensive and efficient climate protection measures. This is the only way to meet the Paris Agreement's goal of keeping global temperature rise below 2° C by 2100. There are extreme consequences that affect the lives of people and animals when temperatures rise above this limit. For example, heat waves, storms, floods, and other extreme weather events occur with increased frequency and intensity.
In order to achieve the 2° C target, a CO_2 tax has already been introduced in various countries as a climate protection measure. But how exactly does such a tax stop climate change? The CO_2 tax sets a price that must be paid by companies and individuals for each ton of CO_2 emitted. The cost thus incurred by the tax motivates emitters of greenhouse gases to reduce the amount emitted. For example, the tax is reflected in higher prices for heating and motor fuels. For this reason, consumers are switching to environmentally friendly alternatives such as heat pumps, electric mobility and the use of renewable energies. The higher the price, the more alternatives prove lucrative. The more environmentally friendly alternatives are used, the greater the amount by which annual CO_2 emissions are reduced.

Based on measurements and data collection, it is known how much CO_2 may still be emitted, until the maximum temperature increase of 2° C is reached. If more CO_2 emissions occur, the world population will have to face the climate change consequences mentioned above. To prevent these, it would be obvious to set the price of the CO_2 tax exorbitantly high. This would force a switch to environmentally friendly alternatives in all areas of application and Ensure a sufficient reduction in CO_2 emissions. This would

Figure 47: Our contribution to climate change: the smoking chimneys of industrial plants. (Source: Pixabay, Pexels)

help the climate, but at what cost? No government, no company, and no private individual will accept to give up its financial well-being for the sake of saving the planet. For this reason, the CO_2 tax issue is not only about meeting the Paris Agreement, but also the avoidance of unnecessary costs. This additional condition presents the difficulty in determining the optimal price for a CO_2 tax. Another difficulty is that not only a constant price for the tax is sought, but a „price path", since the price of the CO_2 tax may vary from year to year.

This results in the following problem: The price path of a CO_2 tax is sought that is set high enough to meet a climate target, such as the aforementioned 2° C target of the Paris Agreement. Further, the price path should be as low as possible to avoid unnecessary costs to businesses and end users. The price path of a CO_2 tax that solves this problem is therefore called the „optimal price path of the CO_2 tax".

The connection between mathematics and stopping climate change does not seem obvious. However, using mathematical methods, it is possible to determine the price path sought for a CO_2 tax. In this type of problem, mathematicians refer to it as an optimization problem. These include maximization and minimization problems, as they are known from school. In this case, it is the latter, because the goal is to minimize the costs incurred by the CO_2 tax. The condition that, in addition, a climate target should be met is an additional restriction on the minimization problem. Through the price of the CO_2 tax, emissions can be managed to meet the restriction. Therefore, determining the optimal price path for a CO_2 tax is, to be more precise, a *optimal control* problem.

Figure 48: Young people protest at a Fridays for Future demonstration. They demand „planet over profit“ (sign inscription). But why choose one? With the help of the CO_2 tax, both are possible. (Source: Markus Spiske, Pexels)

Several mathematical methods exist to solve the given problem. These can be divided into analytical and numerical methods. The analytical solution provides a fundamental underlying idea for the numerical solution and describes the theoretical representation of a price path for the CO_2 tax. Large amounts of data are required for the concrete, realistic calculation of a price path. This includes emissions data from past years in order to make projections for the future. Also relevant are the data representing the impact of environmentally friendly technologies on annual CO_2 emissions. Due to these comprehensive amounts of data, in practice the algorithms of numerical methods are often programmed on the computer. This has the advantage that the scenarios considered are flexible and the data on CO_2 emissions and on environmentally friendly technologies are continuously adaptable. In addition, new relevant factors can be added to the programs at any time. For example, it is possible to incorporate learning factors such as learning-over-time. This takes into account that the technologies that mitigate CO_2 emissions may become less expensive due to exogenous factors, such as general technological progress, become more favorable.

The optimal price path of a CO_2 tax can be calculated in this way for any scenarios. This makes it theoretically possible to halt climate change. In practice, this would require the cooperation of all countries in the world to be able to introduce a global CO_2 tax. However, this is not possible due to geographic, political and social issues. Moreover, the amount of data required for a real and global scenario is too complex and incomplete.

For these reasons, the CO_2 tax has so far only been used in individual countries. Even at the national level, calculations of an optimal price path are based on projections and multi-layered data sets. Nevertheless, the optimal price path provides important indications and serves policy makers as guidance in setting the annual prices of a CO_2 tax.

The field of research around climate change is forward-looking. The successes and findings in this field will determine our lives and those of future generations on this planet. The CO_2 tax is a promising measure to make a positive impact on climate change. It remains exciting to see which countries also decide to introduce such a tax. Due to the topicality of the CO_2 tax issue, the effects and results of this measure can be followed in the coming years.

32 Layout of Large Photovoltaic Power Plants**

The layout of large photovoltaic power plants involves an enormous number of degrees of freedom in the selection, placement and interconnection of the required components. Mathematical optimization drastically reduces the planning effort and increases the quality of the planned plant.

Dr. Petra Bauer works at Siemens Technology, the central department for research and development of Siemens AG, in a team of experts for mathematical optimization. The optimization tool presented here for planning large photovoltaic power plants was developed in a cooperation with Siemens Energy.

The design of large photovoltaic power plants is a complex task that can keep experienced planners busy for weeks. It involves the placement and cabling of the necessary components, taking into account physical laws, external conditions and given target criteria.

A software tool developed at Siemens was able to significantly reduce the planning effort: After input of all data, combinatorial optimization methods calculate a layout in a few minutes. This makes it possible to compare planning variants or change plans without investing days or weeks each time.

Figure 49: Large solar installations consist of hundreds of thousands of solar panels that must be arranged and wired.

Components of a Photovoltaic Power Plant

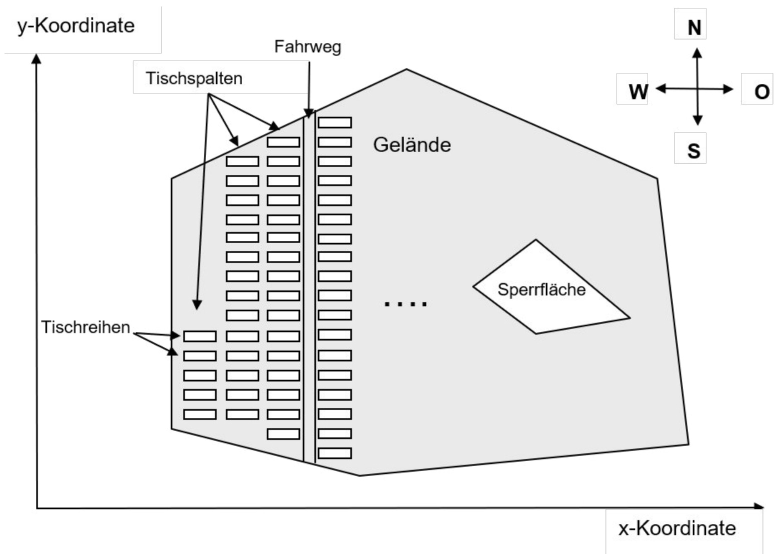
The main components of a photovoltaic power plant are the photovoltaic modules that produce direct current when exposed to sunlight, the tables on which the photovoltaic modules are mounted, inverters that convert the direct current into alternating current, cables that interconnect the modules as well as connect them to the respective assigned inverter, and junction boxes in which several cables are bundled into one thicker cable each.

The Planning Task

The design task is to construct a photovoltaic system on a given site, subject to physical and other restrictions, that maximizes the expected output and does so as cost-effectively as possible.

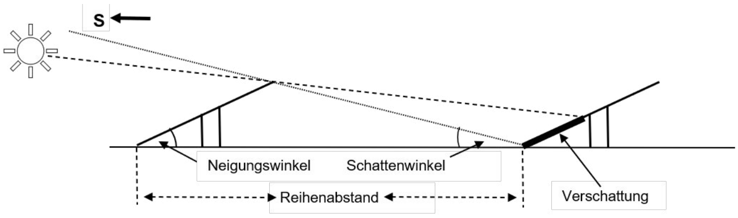
To illustrate the complexity of the problem, we describe some aspects of the task in a bit more detail below.

The terrain is given by a polygon course describing the outline, and, if necessary, by further polygons for restricted areas, which must not be built on. For the approximation of the terrain topography, we also need elevation data at a sufficient number of locations as a basis for triangulation, i.e. an approximation of the terrain by plane triangles.



The solar modules are placed, inclined to the south, in rows running from east to west. If the terrain slopes to the north, the distance to the next row of tables (as seen from the south) must be greater due to shading effects, than if the terrain rises to the north.

One would now like to place as many tables as possible under the constraint, that routes from south to north exist according to parameterizable defaults and at a likewise adjustable



time of the year (e.g. on the longest day of the year at 12 o'clock noon) the modules do not shade each other. The rows must be located between two driveways at the same „height" in the south-north direction to be easily accessible for maintenance.

In addition, inverters must be selected and placed that together have sufficient capacity to convert the direct current produced into alternating current. Each table must also be assigned to an inverter and connected to it in a hierarchical cable structure.

When wiring, it is important to ensure that it is always perpendicular and that all cables, that run in a south-north direction must be laid underground, which incurs additional costs. Cable diameters must be chosen in the best possible way in consideration of acquisition costs and losses due to cable resistance.

The Optimization Tool

This overall task is too complex to be solved by an optimization procedure, which makes all decisions at the same time. Here is a list of the most important decisions to be made:

- location of the guideways
- location of the tables in rows and columns
- selection of inverters
- location of inverters and assignment of tables to inverters
- cable routing taking into account excavation costs
- Determination of cable diameters

We rely on decomposition of the overall problem and solve individual tasks, although not independent of each other, one by one. In order to keep the associated sacrifices in the solution quality as low as possible, we consider in the individual substeps, as far as possible, the aspects of the following steps are taken into account. For example, when determining the exact location of the routes, an approximation to the cabling costs is already included in the objective function.

Integer Linear Programming, Dynamic Programming, and heuristics are used in solving the subproblems. In Integer Linear Programming, the restrictions and the objective function are modeled by linear equations and inequalities in integer variables and solved with appropriate solvers, in Dynamic Programming, the solution of a problem is traced back to the solutions to subproblems, which are stored and then suitably combined. When problems are too complex to be solved exactly fast enough, one applies heuristics, which compute feasible solutions by „skillful “approach.

The optimization tool we developed allows the planner to import terrain data on the user interface, specify the available components (module types, tables, inverter types, cable types, etc.) and formulate constraints.

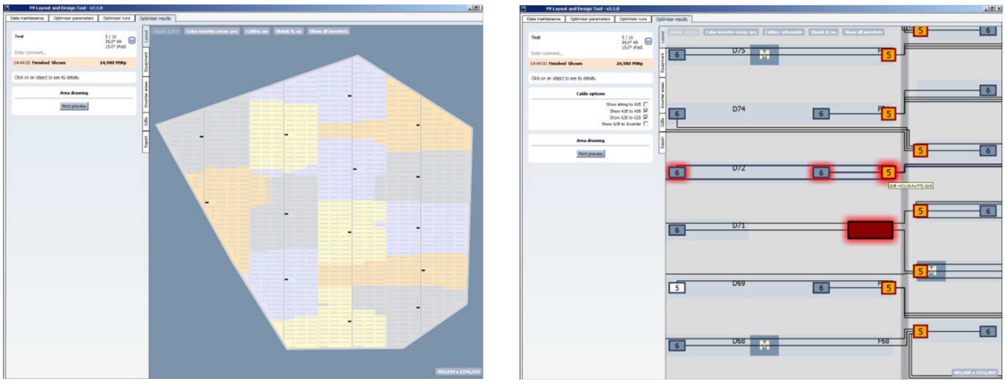


Figure 50: 25 MWp power plant with tables (gray), inverters (black), inverter areas (colored) and paths. On the right a detailed view.

At the push of a button, a layout is calculated in a few minutes by a hierarchy of mathematical optimization methods, evaluated and displayed on the surface. By zooming into the layout, details can be viewed and modified by the planner as desired. A subsequent reoptimization step that takes the planner’s wishes into account can also be performed.

33 Optimized Gas Heater Design**

Or: How do I reach the desired temperature quickly and comfortably? Mathematical optimization improves comfort and efficiency.

Matthias Stursberg, R&D Manager at Vaillant GmbH in Remscheid reports together with Tobias Suszka and Prof. Dr. Kathrin Klamroth from the University of Wuppertal about their joint project.

As in almost all technological fields, the development of gas heating appliances is becoming increasingly complex and elaborate. Numerical simulation and mathematical optimization have numerous applications here, because even small increases in efficiency can lead to quite significant CO_2 savings. However, the comfort of water heaters should not be neglected either, because after all, we don't want to wait long for the hot water to come out of the shower.

The key to this lies in the control of water heating, so, for example, of instantaneous water heaters for the preparation of hot water for showers and sinks. Ideally, the heater will provide a constant outlet temperature of, say, 55°C, regardless of the amount of water passing through it. This is then mixed to the desired temperature using the shower or faucet mixer. On the one hand, slow changes in the outlet temperature over time should be avoided, otherwise the temperature must be permanently readjusted with the mixer when showering or washing hands. On the other hand, a changing water flow causes so-called temperature overshoots, which impair comfort: The water briefly becomes much too hot – so that you have to jump away – and immediately afterwards it is too cold again. The goal of an optimization is now to determine a suitable compromise with the help of an analysis of the trade-off between the two criteria „too fast" and „too slow" regulation of the outlet temperature – just in such a way that the user neither has to permanently readjust nor has to struggle with strong temperature jumps.

For this purpose, *evolutionary multicriteria optimization methods* (EMO) are used. These mimic the evolutionary process known from biology. A population of solutions is generated, which is transformed into a new population by selection, mutation and recombination. This successor generation should contain better solutions because of selection pressure. The process continues for several generations until a satisfactory result is found or no more improvements are made. The selection, i.e. the evaluation of a solution, is carried out using simulation models. The behavior of the respective device under representative load cycles is predicted. In this way, an approximation of the so-called Pareto Front is obtained. The Pareto Front contains all solutions that cannot be improved simultaneously with respect to both objectives. I.e. an improvement in one objective necessarily causes a deterioration in the other objective.

Figure 51 shows the result of such an optimization based on a generic simulation model. The lower part of the graph shows the approximated Pareto front. Each point corresponds to a solution. For each solution, the value of the total deviation and the value of the overshoot are contrasted. The total deviation is the deviation from the desired temperature integrated over time, and the overshoot is the largest deviation from the desired temperature. Both should be as small as possible: Both the total deviation from the desired temperature and the overshoot. It can be seen, however, that both cannot be achieved at the same time – the perfect solution is not possible and we must find a suitable compromise. More precisely, there is a trade-off here between the two optimization goals: If we want to reduce the total deviation, we can only do so at the expense of increasing overshoot. As an example, for three solutions (A, B and C) the corresponding temperature curves with two load changes each are also shown, showing the respective advantages and disadvantages.

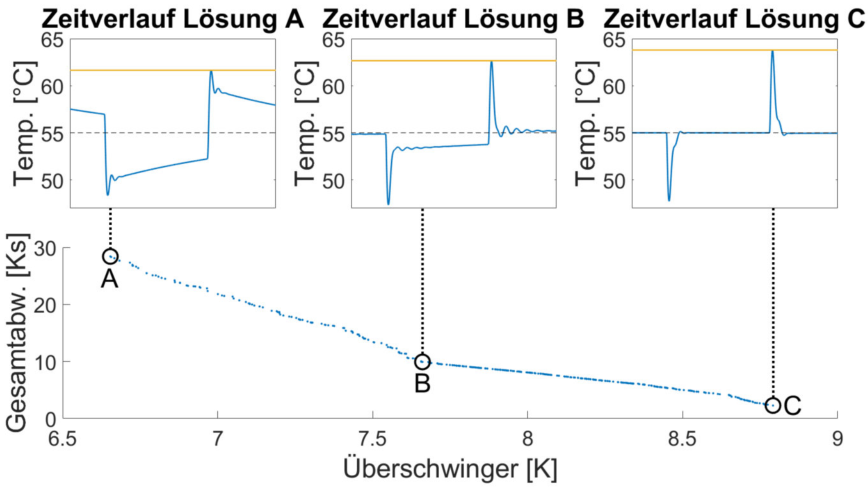


Figure 51: Alternative controls of an instantaneous water heater and a good compromise between fast and comfortable water heating.

Optimization problems naturally also occur in many other areas of heater development. Within the scope of jointly supervised theses and doctoral dissertations, current issues have been and are being addressed using modern methods adapted to the respective problem. For example, the parameters of simulation models were optimized within the scope of master's theses in order to adapt the quality and predictive power of the simulations as well as possible to the data measured in the laboratory (Figure 52). In the future, AI methods will also be increasingly used, e.g., when it comes to optimized sequence planning of measurement cycles.

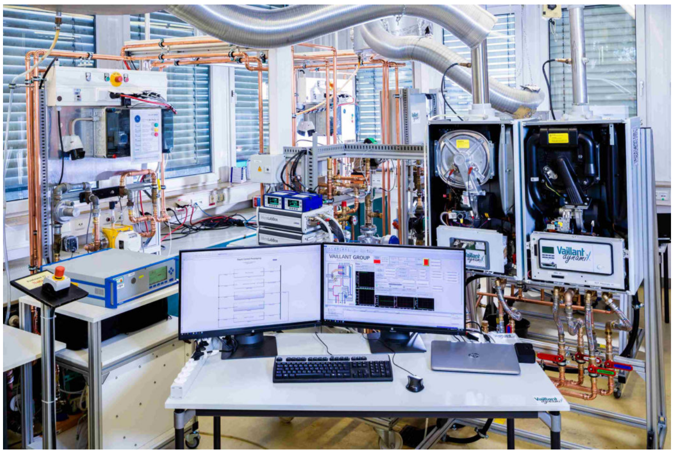


Figure 52: A measuring stand in the laboratory of Vaillant GmbH (Photo: Mike König, Remscheid).

34 Statistics Determine the Choice of Wind Turbine*

Which type of wind turbine will provide the most energy at which location? To do this, you need to know what the odds are for all wind speeds at that location. The Weibull function provides this chance.

Nothing is as variable as the wind. Sometimes it storms, sometimes it is calm. Sometimes the wind blows from the west, sometimes from the east. But wherever you measure wind speed on Earth, the chance of a given wind speed is given by a mathematical function, the Weibull function. Swedish engineer and scientist Waloddi Weibull first described this function in 1939.

The Weibull function differs from the best known distribution from statistics, the standard or normal distribution. The standard distribution is symmetric; the Weibull function is asymmetric. It is actually a standard distribution sampled at positive wind speeds. For the Weibull function, only the magnitude of the velocity matters, not the direction. As with a wind turbine, the wind direction does not matter because the turbine rotates with the wind.

In the Weibull function, the probability of zero wind speed is very small. As the wind speed increases, the probability initially increases up to a certain maximum. After that, the chance of increasing wind speed slowly decreases. Only two factors determine how stretched or peaked the probability distribution is: A scaling factor and a shape factor.

There are many different wind speeds in the Netherlands, which means, that the Weibull function is widely spread over the whole country. On the other hand, there are relatively few different wind speeds at the equator: There, it either blows hard or not. This leads to a strongly pronounced Weibull function. When installing a wind turbine or wind farm, it is essential to know exactly the Weibull function at the intended location. This is because it makes it possible to calculate the expected energy yield.

Communicating Vessels

„To determine the Weibull function at a particular location, you need to measure the wind at that location continuously for at least a year“, says wind energy professor Gerard van Bussel of Delft University of Technology. „At least you have experienced all seasons once. You then need to combine the information from the measured Weibull function with the turbine manufacturer-supplied information on the relationship between turbine power and wind speed for a given type of wind turbine. From this, you can calculate which wind turbine is the best fit for that location and the estimated energy yield.“

At first, you might think that you should choose a wind turbine that delivers as much energy as possible. In practice, however, energy companies are more interested in providing as reliable a supply of a particular energy as possible. „Unfortunately, maximum energy output and maximum reliability are two communicating vessels“, says van Bussel, „and so you have to choose between the two. Then you may produce the most kilowatt-hours with a five-megawatt system, but with a two and a half megawatt plant, you're delivering the most reliable energy, simply because those two and a half megawatts are reached sooner.“

In 2014, the Netherlands will generate 4.5% of its energy from wind power, or 2,300 megawatts. The plan for 2020 is for no less than twelve thousand megawatts to come from wind, half of which will come from onshore wind farms and the other half from offshore wind farms.

Windmills at Sea

The Netherlands now has two offshore wind farms: North Sea Wind in Egmond aan Zee and the Prinses Amalia offshore wind farm in IJmuiden. Three more are to be added by 2020. „We are still learning how to build offshore wind farms,“ says van Bussel. „It's not wise to use land-based turbines offshore as well. Wind turbines wear out much faster at sea, partly because the air is so salty and the wind blows harder. In addition, repairs at sea are expensive and far from possible every day. That requires robust turbines.“

Many people underestimate the demands placed on wind turbines, van Bussel says. „The largest turbine in the world has rotor blades one hundred and eighty meters in diameter. That's two and a half times as long as the world's largest passenger plane, the A380. And this rotor diameter even goes up to two hundred and fifty meters.“ What's more, a wind turbine has to operate continuously for twenty years under conditions comparable to the take-off and landing conditions of an airplane, in other words, at the very moments when an airplane suffers the most. Van Bussel: „In twenty years, wind turbines will look the same, but they will all have clever tricks, just as cars and airplanes have evolved from low-tech to high-tech in recent decades. But the Weibull function will remain unchanged.„

Figure 53: Example of Weibull functions at different form factors (k).

35 Immigration and Integration in Changing Societies***

A mathematical network model simulates immigration and coexistence processes in the European Union.

Dr. Yao-Li Chuang and Prof. Dr. Maria R. D'Orsogna, Department of Mathematics, California State University, Los Angeles, summarize their work on immigration modeling:

Human migration has always been a part of human history. Our earliest ancestors left Africa over 70,000 years ago and settled in Europe, Asia and the Americas in search of new hunting grounds. By 10,000 B.C.E., most of Earth's landmass was settled, and humanity adapted to new environments, learned to cooperate, developed new tools, and invented agriculture.

More recently, workers leave their homelands in search of jobs or better economic opportunities; refugees flee civil wars, brutal dictatorships, or religious persecution. Others migrate after natural disasters, in response to famine, drought, or loss of arable land. According to the International Organization for Migration (IOM), international migrants account for 3.5% of the world's population in 2019, a record high. Climate change is expected to displace many more people worldwide.

The United States has experienced many waves of immigration since its founding. Pilgrims and Puritans were among the first to settle on the East Coast in the 17th century, seeking religious freedom. Enslaved West Africans came against their will for more than a hundred years until 1865, when the 13th Amendment abolished slavery. In the 19th century, a massive famine drove many impoverished Irish to the East Coast. Large numbers of Germans arrived during the same period to escape poverty and unrest following the March Revolution and the 1848-1849 uprisings. They settled in the Midwest and established thriving communities. In the 2000 census, Americans of German descent outnumbered any other group. The California Gold Rush attracted Asian immigrants, especially from China, until the 'Chinese Exclusion Act' was signed in 1882. Finally, large-scale immigration from southern and central Europe occurred in the early 20th century, mainly destitute Italians and persecuted Eastern European Jews. The Great Depression of 1929 and World War II halted the influx of migrants to the United States, which resumed during the Cold War with the immigration of political refugees from the USSR and Cuba. In 1965, the 'Immigration and Nationality Act' was passed, eliminating quotas based on nationality and allowing family reunification. This law, along with significantly improved economic conditions in Europe, changed migration patterns to the United States so that today the majority of new arrivals are from Asia and Latin America. Europe, a source of migrants toward the United States for more than three centuries, has itself begun to attract migrants from Africa, Asia, and the Middle East in recent years.

Intracontinental migration between continents is also on the rise, aided by EU legislation. A particularly vivid example of large-scale migration to Europe is the arrival in

Germany of more than a million people, and on the rest of the continent of more than a million people, mostly refugees from Syria, beginning in the summer of 2015 and following the civil war that broke out in Damascus in 2011. Much has been written about that fateful summer: the remarkable generosity of the German people; the political differences at the national and European levels over the right refugee policy; the strain on EU border countries such as Italy, Greece and Hungary, which did not have the resources or the will to cope with such large numbers of new arrivals; the exploitation of migrants by smugglers; and the gradual rise of discontent in much of society over real or perceived issues of security, identity, secularism and multiculturalism.

Regardless of the larger sociopolitical discourse, one truth remains: Many migrants have settled in European cities in recent years and will continue to do so, similar to the United States, which has a longer immigration tradition and history of native rejections, success stories, and societal changes brought about by migrants' delicate transition from outsiders to insiders.

So a key question arises: how can we ensure that the arrival of newcomers is beneficial to both migrants and long-term residents? Integrating immigrants into the broader society is essential to creating vibrant and cohesive communities. Newcomers who do not integrate well due to circumstances, native reluctance, lack of motivation, and/or resources tend to self-segregate and form island groups. While these enclaves offer immigrants advantages and a sense of belonging, they can also prevent them from full civic participation and lead to the risk of creating parallel societies.

Many mathematical models have been introduced to study migration, which is often described as a *'push-pull phenomenon'*: The push arises from poor socioeconomic conditions in the migrants' home countries, and the pull results from more desirable conditions in the destination countries. One of the first scholars to describe migration in push-pull terms was the Anglo-German geographer Ernst Ravenstein, who in the late 1800s established a set of „laws of migration“ upon which modern migration theory is built. Later, these laws were mathematized, resulting in various expressions for M_{ij}, the number of persons, moving from location i to location j. A general form is of the type

$$M_{ij} = k\,\frac{R_i + A_j}{d_{ij}},$$

where k is a proportionality constant, d_ij is the distance between origin i and destination j, and R_i, A_j are the „repelling“ and „attracting“ properties of origin i and of destination j, respectively. Examples of R_i, A_j might be unemployment rates, wages, housing costs, or people in the labor market competing for the same jobs. An alternative form is of the type

$$M_{ij} = k\,\frac{P_i^{\alpha} + P_j^{\beta}}{d_{ij}^{\gamma}}$$

where P_i, P_j are the populations in the origin i and destination j areas, and α, β, γ are parameters to be fitted to the data. Net migration from region i is given by

$\sum_j (M_{ij} - M_{ji})$, where is the sum over all possible jurisdictions j, to or from which there exist immigration or emigration flows connected to territory i. Sometimes the known migration flows are used to reconstruct estimates for the push and pull factors R_i, A_j, other times the data are used to estimate the values of α, β, γ. These models have been applied over several years to study international and internal migration, including within and across the U.S., U.K., New Zealand, Italy, Vietnam, between the EU and neighboring countries, and have been further adapted to include precipitation and temperature changes to study the effects of climate change on migration.

To better understand problems related to immigrant integration, a mathematical model, that examines societal dynamics by incorporating elements from network theory and game theory can be used. Society is assumed to consist of two groups, N_h „hosts“ and N_g „guests“, represented as nodes interacting in a time-varying social network. Each node represents an individual or a small group of like-minded individuals, either host or guest. Both types interact with each other based on rational choices as they seek to improve their socioeconomic status, which leads to changes in the social network. Each node i carries a time-dependent attitude x_i^t toward others and is given a utility function U_i^t to maximize, which depends on its m_i^t connections. Over time, to increase their utility, nodes adjust their attitudes and create new or cut old connections; as a result, the network evolves toward integration or segregation between hosts and guests. While the utility function follows game-theoretic rules, the attitudes are assumed to evolve through opinion dynamics. The two influence each other in a synergistic way. For an illustration, see Figure 54.

The settings x_i^t vary between $0 \le x_i^t \le 1$ for hosts and between $-1 \le x_i^t \le 0$ for guests; the amount $|x_i^t|$ indicates the degree of hostility toward the other group. Thus, $x_i^t \to 0^{\pm}$ characterizes the most open-minded guests or the most hospitable hosts, while $x_i^t = \pm 1$ represents the highest level of xenophobia. Utility U_i^t is given by a pairwise reward to which each node j connected to i contributes, and by a cost function for maintaining m_i^t connections such that

$$U_i^t = \sum_{j \in \Omega_i^t} A_{ij} \, \exp\left(-\frac{|x_i^t - x_j^t|^2}{2\sigma}\right) - exp\left(\frac{m_i^t}{\alpha}\right).$$

Here Ω_i^t is the set of nodes, that are connected to i at time t, so m_i^t is given by its cardinality, $|m_i^t = \Omega_i^t|$. The pairwise reward is the first term on the right-hand side of the above equation and depends on the attitude difference $|x_i^t - x_j^t|$ between nodes i and j. The smaller the attitude difference, the higher the reward. Thus, if both i and j belong to the same group, both hosts and guests, the reward for $x_i^t = x_j^t$ is maximized, leading to consensus between them. However, if i and j are from different groups, the reward is optimized only when both nodes adopt a more cooperative stance, $x_i^t \to 0^-$ and $x_j^t \to 0^+$. The parameter σ controls the sensitivity of the reward, and the amplitude A_{ij} specifies the maximum achievable reward. For simplicity, A_{ij} can be set as a constant. The cost function is the second term on the right-hand side from the above equation and decreases with the number of connections m_i^t, modulated by the scaling coefficient

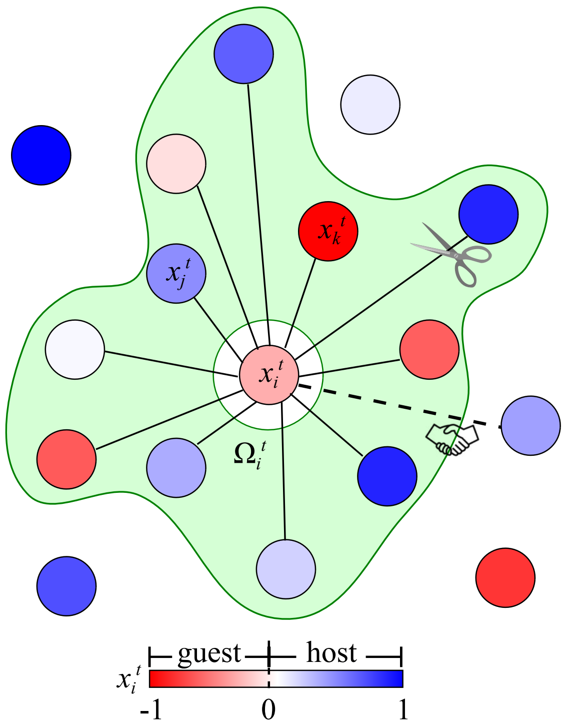


Figure 54: Each node i is characterized by a variable setting $-1 \leq x_i^t \leq 1$ at time t. Negative (red) values represent guests, positive (blue) values represent hosts. The amount $|x_i^t|$ represents the degree of hostility of node i towards members of the other group. All nodes j, k connected to node i represent the green shaded social circle Ω_i^t of node i at time t. The utility U_i^t of node i depends on its attitude relative to that of its m_i^t connections. Nodes maximize their utility by adjusting their settings x_i^t and making or breaking connections.

α. This term limits the number of active connections that node i can establish.

As we progress from time step t to time step $t+1$, the connectivity of a given node i is modified to maximize utility U_i^t; its setting x_i^t is changed by imitation. In practice, a random node j is selected and its relationship with node i is evaluated. If nodes i and j are connected and disconnecting them increases the utility for node i, then the connection is disconnected, otherwise the two nodes remain connected; if the two nodes are not connected and making a connection between them increases the utility for node i, the connection is made, otherwise the two remain disconnected. Settings are changed by selecting another random node ℓ that is connected to i, and by decreasing the gap between x_i^t and x_ℓ^t. The idea is that once a connection is made, the two nodes tend to share closer views. For example, if node i is a host, then

$$x_i^{t+1} = \max\left(0, x_i^t + \frac{x_\ell^t - x_i^t}{\kappa}\right) \quad \text{for hosts.}$$

Here, the setting of node i at time $t+1$ is closer to the setting of node ℓ that node i is trying to mimic, but we keep the lower bound at 0 because by construction all host nodes must have a positive setting. Similarly, if node i is a guest, then we have

$$x_i^{t+1} = \min\left(0, x_i^t + \frac{x_\ell^t - x_i^t}{\kappa}\right) \quad \text{for guests.}$$

The parameter κ in both equations governs behavioral adaptation: small values of κ represent fast imitation between nodes i and ℓ, while large values of κ represent slower imitation.

Finally, the initial conditions model the way hosts are initially integrated (or not) into the community. An extreme case is that of a perfectly executed „welcome" program, where guests have sufficient social ties to hosts and all nodes are randomly connected, regardless of their attitude and utility. This is the initial condition used in Figure 55. The other initial condition is that of nonexistent initial resources, where the guests arrive in a completely foreign environment. The hosts are naturally connected, in their own equilibrium state, and the guests are introduced without any connections to hosts or other guests.

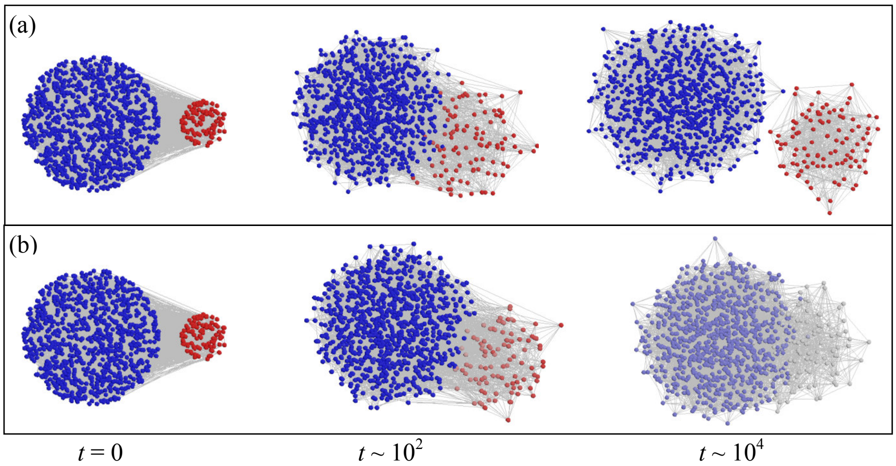


Figure 55: Simulated dynamics leading to (a) full segregation and (b) integration between guest (red) and host (blue) populations. Initial conditions are randomly connected guest and host nodes with the settings $x^0_{i,\text{guest}} = -1$ and $x^0_{i,\text{host}} = 1$. The two figures differ only by κ, the time scale of the setting adjustment, with $\kappa = 1000$ in figure (a), where segregated clusters emerge, and $\kappa = 100$ in figure (b), where a connected host-guest cluster emerges over time.

In Figure 55 we show two representative steady-state results starting from the „perfect" initial conditions of mixing hosts and guests at the beginning. The only difference between the two panels is the parameter κ. In the top panel, where the parameter κ is large, despite the most favorable initial conditions, hosts and guests are separated and behave in an extremely hostile manner. Any initial imposed cross-group utility results in low rewards, which do not increase over time so that all ties between hosts and guests are eventually cut. Accumulation of socioeconomic wealth tends to occur through insular intra-group linkages, so that eventually enclaves are created, where the two separate groups adopt uniform but different attitudes x_i. In the lower panel, where the value is low, all nodes develop more cooperative attitudes that the cross-group rewards, so that hosts and guests remain mixed. Finally, $x_i \to 0$ on all nodes.

Other factors that may lead to enclave formation include a high guest-guest ratio (N_g/N_h), which increases the likelihood of intra-group connections and reduces communication between the host and guest populations, and low values of amplitude A_{ij}, which yield low socioeconomic gains for host-guest connections. Thus, one possible way to avoid segregation is to actively promote host-guest interactions, e.g. by providing incentives to improve the cultural fit of guests and hosts (lowering the factor in our model), ensuring that local communities are not overwhelmed by newcomers (lowering the local guest-host ratio N_g/N_h), giving preference to guests who bring desired skills, or helping them acquire these skills (increasing the amplitude A_{ij}).

Note that cultural adaptation does not necessarily mean that both sides must give up their identities, but that to promote integration, the different groups must accept each other and try to build a relationship and bridge their differences. This, of course, is the long-term challenge for the future. For more details, we refer the interested reader to. [33].

[33] Y.-L. Chuang, T. Chou, M.R. D'Orsogna, *A network model of immigration: Enclave formation vs. cultural integration*, Networks & Heterogeneous Media 14(1) (2019), 53-77. DOI: 10.3934/nhm.2019004

36 Mathematical Modeling of Radicalization Processes***

Extreme opinions spread like contagious diseases. And this is exactly how one can mathematically model the spread and radicalization.

Dr. Yao-Li Chuang and Prof. Dr. Maria R. D'Orsogna, Department of Mathematics, California State University, Los Angeles, summarize their work on modeling radicalization processes:

The rapid spread of radical ideologies has led to a worldwide succession of terrorist attacks in recent years. Understanding how extremist tendencies emerge, develop, and drive individuals to act is important from a cultural perspective, but also to formulate response and prevention strategies. Demographic studies, interviews with radicalized individuals, and analysis of terror databases show that the path to radicalization occurs in multiple stages, with age, social context, and peer exchanges playing important roles.

Mathematical modelers have used tools from statistical mechanics and applied mathematics to describe existing and new explanatory models and to propose new counter-terrorism strategies. Here we will introduce the compartmental modeling approach for populations with increasingly extreme views.

The seeds of pre-radicalization are often rooted in various unaddressed grievances and personal frustrations, such as lack of employment and opportunity, racial and religious discrimination, and social exclusion. Individual discomfort can also come from socioeconomic injustices, opposition to policies perceived as too progressive, or conversely, from a desire for profound social change. Personal discontent leads to self-identification, in which marginalized or self-seeking individuals gradually begin to construct new identities and routines, and turn away from old ones.

Like-minded people are actively sought out and new friendships are formed. Mutual encouragement and the lack of a counter-dialectic allow extremist views to reinforce themselves and become entrenched. Once an extreme view has been adopted, it becomes difficult to deny the new ideals; justification and praise of violence follow. Another activism phase follows, when radicals commit to militantly spreading their beliefs to others, until external events such as political or judicial decisions or simple coincidences crystallize a willingness to act violently. A schematic representation of this process is shown in Figure 56.

Radicalization typically occurs through networks of peers and can be facilitated by technologies such as web-based recruitment materials or chat rooms; IS is a well-known Internet-savvy group that uses social media to recruit Western foreign fighters in Syria and Iraq. A RAND study[34] shows that the virtual world allows extremists to communicate, collaborate, and persuade without physical contact, and to connect with like-minded people from around the world.

[34] I. von Behr, A. Reding, C. Edwards, L. Gribbon, *Radicalization in the digital era: The use of the internet in 15 cases of terrorism and extremism*, RAND Europe (2013).

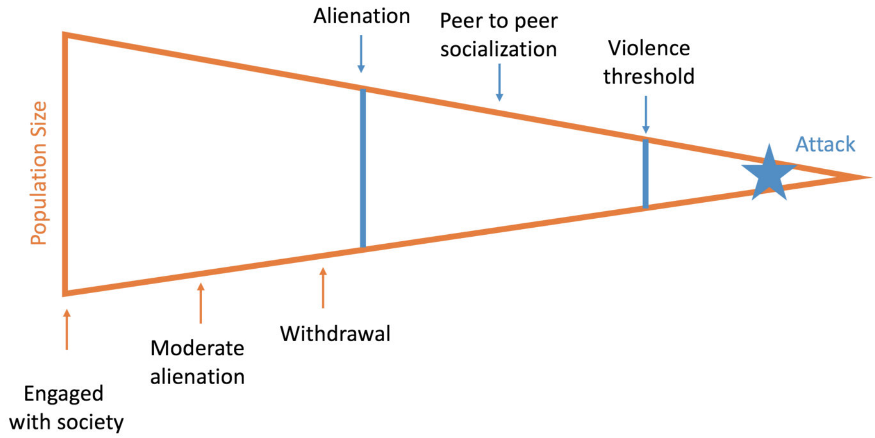


Figure 56: Schematic representation of the radicalization process. Committed individuals go through progressive phases of withdrawal, until they separate themselves from the rest of society. From then on, they cultivate a new identity, seek out like-minded people and prepare for violence. The process culminates in possible terrorist attacks. Not all individuals go through the entire hierarchy: The number of radicalizing individuals decreases as the level of extremism increases, resulting in a multilevel horizontal funnel.

Although still in their infancy and mostly experimental, de-radicalization programs are being developed and implemented around the world with great expectations. They include education, exposure to literature, sports and the arts, psychological counseling, job training and informal one-on-one meetings. The goal is to prevent radicalization, release known violent extremists, and reintegrate former convicted extremists into the community.

Radicalization in Opinion Dynamics Models

We discuss here radicalization models embedded in the context of opinion formation, a very natural starting point since radicals can be viewed as individuals with extreme „opinions". Here, an individual's beliefs – and those of society as a whole – change through personal contact, the influence of the media, or current events. In this context, radical opinion and its spread are modeled similarly to an epidemic (cf. Chapter 10).

Mean-field models assume that individuals with a particular belief behave uniformly and define a homogeneous population; each of these populations is assumed to interact with others, affecting their views in reversible or irreversible ways, such that over time the size of a given population may increase or decrease. These dynamics are usually described by systems of ordinary differential equations (DGLn). These simplified models do not account for the mechanisms of how individuals are predisposed to fanaticism, but they do allow for the identification of parameters or mechanisms that drive observed trends. The model of Carlos Castillo-Chavez and Baojun Song[35] is one of the first models to introduce population dynamics into the study of the transmission of fanatic behaviors.

Concepts such as (effective) reproduction numbers R_0, which are used to predict when a disease becomes endemic ($R_0 > 1$) and when it does not ($R_0 < 1$), are adapted to the spread of radical ideologies. In particular, a so-called „fanatic hierarchy" is introduced with the total population $T(t)$ divided into a subgroup $G(t)$ that has no propensity to radicalize and three subgroups of individuals that are at different stages of their commitment to the extreme ideology. The latter are $S(t)$, a group of susceptibles who are not yet radicalized but are susceptible and open to the radical ideology in question; the population $E(t)$, individuals who have recently turned into fanatics („semi-fanatics"); and finally $F(t)$ individuals who are fully committed to radical views. The sum of the different subpopulations gives the total population: $T = G + S + E + F$. The model of Castillo-Chavez and Song is as follows

$$\begin{aligned}
\frac{dG}{dt} &= \Lambda - \beta_1 \frac{G\,C}{T} + \gamma_1 S + \gamma_2 E + \gamma_3 F - \mu G, \\
\frac{dS}{dt} &= \beta_1 \frac{G\,C}{T} - \beta_2 \frac{S\,(E+F)}{C} - \gamma_1 S - \mu S, \\
\frac{dE}{dt} &= \beta_2 \frac{S\,(E+F)}{C} - \beta_3 \frac{E\,F}{C} - \gamma_2 E - \mu E, \\
\frac{dF}{dt} &= \beta_3 \frac{E\,F}{C} - \gamma_3 F - \mu F,
\end{aligned}$$

where $C(t) = T(t) - G(t)$ represents those individuals that are susceptible, partially radicalized, or fully radicalized („core population"). The system includes an entry rate, called „birth rate", Λ into the non-susceptible population G and a universal exit rate („death rate") μ for all subpopulations, such that after summing all four equations $dT/dt = \Lambda - \mu T$, which yields $\lim_{t\to\infty} T(t) = \Lambda/\mu$. All other terms are associated with specific transitions between the different subpopulations, namely recruitment rates β_i and return rates γ_i. We replace T by the limit Λ/μ and $G(t)$ by $\Lambda/\mu - C(t)$ in the above DGL system for the state variables S, E and F. The reduced system is „dynamically" equivalent.

A second global threshold $R_3 = \beta_3/\gamma_3$ controls the establishment of the fanatic population $F(t)$ and consequently the persistence of the fanatic ideology. If $R_3 \leq 1$ holds, then the fanatic population dies out in the limit: $\lim_{t\to\infty} F(t) = 0$. For this case $R_3 \leq 1$ one uses this limit and the dimension of the model can be further reduced. One obtains the following two-dimensional system:

$$\begin{aligned}
\frac{dS}{dt} &= \beta_1 (1 - C) C - \beta_2 \frac{S\,E}{C} - \gamma_1 S, \\
\frac{dE}{dt} &= \beta_2 \frac{S\,E}{C} - \gamma_2 E, \\
C &= S + E.
\end{aligned}$$

[35] C. Castillo-Chavez, B. Song, *Models for the transmission dynamics of fanatic behaviors*, Bioterrorism-Mathematical Modeling Applications in Homeland Security, Philadelphia: SIAM 28 (2003), 155-172.

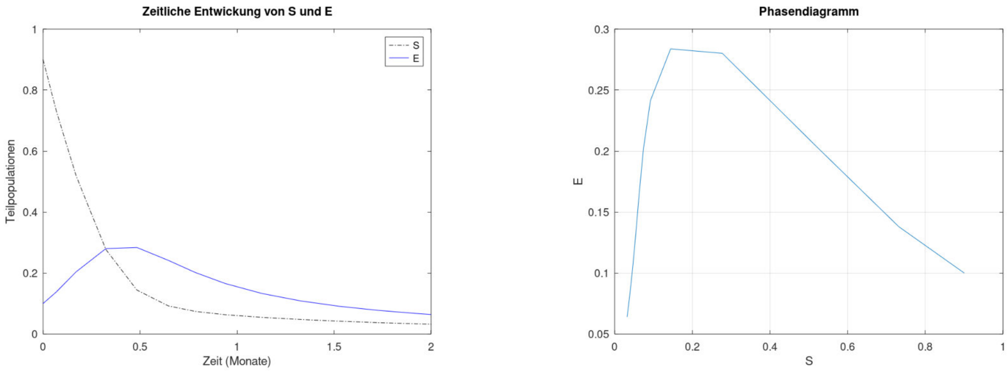


Figure 57: Simulation of the reduced two-dimensional system.

In the Figure 57, one sees a simulation of this system with parameters $S(0) = 0.9$, $E(0) = 0.1$, $\beta_1 = 3$, $\gamma_1 = 2$, $\beta_2 = 10$, and $\gamma_2 = 4$. Notice how the semi-fanatic subpopulation $E(t)$ first increases to 0.28 and then decreases.

For example, individuals in the non-receptive population G may become receptive to radical ideologies when exposed to them; the push of the radical message is modeled by the population ratio $C/T < 1$ and the associated rate β_1, so that the net flow $G \to S$ is expressed by $\beta_1 GC/T$. Similarly, the flow from the prone state to the first stages of fanaticism $S \to E$ depends on indoctrination and example from the more radical individuals and is given by $(E + F)/C$ modulated by rate β_2. The transition from the initial radical state to the fully engaged state $E \to F$ is instead driven solely by the radical group via F/C and modulated by rate β_3.

The ODE system defines a hierarchy in the sense that transitions through the $G \to S \to E \to F$ stages are driven by more radical populations affecting the weaker ones. The three feedback rates γ_i, for $i = 1, 2, 3$, represent de-radicalization from the subgroups of susceptibles, recent converts, and fully fanaticized, respectively. There are no reverse intermediate transitions in the model: Radical individuals F can, for example, return to the non-susceptible population G, but they do not return to their first radicalization stage E.

Despite its simplicity, the model provides useful insights into social outcomes, in terms of attractor points and thresholds. Of course, the most important question is under what conditions a finite fanatic population F emerges. Analysis of the DGL system shows that for $\gamma_1 > \beta_1$ no level of radicalization is maintained ($S = E = F = 0$), implying that one way to avoid radical discourse in a society is to prevent the radical process from beginning in the early $G \to S$ stage, when individuals become susceptible to extremism.

Depending on other parameter combinations, attractor points can be identified without fanatic populations ($S^* \neq 0, E = F = 0$) or ($S^* \neq 0, E^* \neq 0, F = 0$), and with fanatic populations ($S^* \neq 0, E^* \neq 0, F^* \neq 0$). Equally important are the initial conditions and time scales: For example, even if is predicted to be at equilibrium $F = 0$, an initially small group of extremists may successfully infiltrate the population, leading to a large fanatical population before it begins to decay.

Numerous other studies are based on this DGL system with four variables. For example, the possibility of two competing radical groups starting from the same general nonprone population G was modeled by including two recruitment rates $q\beta_1$ and $(1-q)\beta_1$ that result in two different prone groups. Each of these two core groups follows the same radicalization process shown in the DGL system without interacting. Both subgroups also cross-contact through the $S \to E \to F$ hierarchy. There is competition in recruitment and retention efforts between the two radical branches, so that fanatics gain new followers from converts of the opposite ideology. In particular, it is shown that the two groups cannot both coexist in equilibrium without mutual interactions: competitive recruitment is necessary for the emergence of two finite fanatic populations F_1, F_2.

Compartmental models such as this one have also been applied to current extremist movements, for example, the influence of separatist groups in the Spanish Basque Country[36]. Subpopulations were created and transition parameters were estimated using demographic and election data; immigration and emigration data were also included. The ideological evolution of society was thus analyzed over a 35-year period, including projections into the future. Similarly, this ODE system has been adapted to study the radicalization of far-right movements in Germany[37] where surveys on the *Likert scale*, repeated every two years starting in 2002, were used for calibration. Users were found to go through several levels of engagement until they become recruiters or real terrorist actors. Alternative compartmental models, for example, define subpopulations by cultural background or simply by age, for example.

For more details, we refer the reader to the extensive work by Yao-Li Chuang and Maria R. D'Orsogna[38], which also presents alternative mathematical models of radicalization processes.

[36]F.J. Santonja, A.C. Tarazona, R.J. Villanueva, *A mathematical model of the pressure of an extreme ideology on a society*, Computers and Mathematics with Applications 56 (2008), 836-846.

[37]T. Deutsch, *Mathematische Modellierung von Radikalisierungsprozessen am Beispiel von rechtsradikalen Gruppierungen in Deutschland*, Bachelor thesis, Bergische Universität Wuppertal, November 2014.

[38]Y.L. Chuang, M.R. D'Orsogna, *Mathematical models of radicalization and terrorism*, arXiv preprint (2019):1903.08485.

37 Fair School Place Allocation: From Boston to Around the World**

How do you actually find good assignments? Which student gets a spot at her school of choice? Which student gets his or her elective? Which teams should compete in tournaments? And when is an assignment actually good?

Dr. ’Agnes Cseh from the Hasso-Plattner-Institute, Potsdam and Prof. Dr. Nicole Megow from the University of Bremen report on how to find good mappings:

Good and fair assignments are needed all the time and in all different walks of life. Which student gets a spot at her school of choice? How are students assigned to their courses? Which applicant will be admitted to which university? Which teams will compete in tournaments? How will dorm spots be allocated? What all these questions have in common is that a limited amount of resources are to be allocated to applicants, and in doing so, different preferences must be taken into account.

Mathematicians are concerned with the question, how to find a good allocation. For this purpose, algorithms are developed, i.e., procedures that can find such an assignment. But what is a good assignment? Is it possible to measure and compare goodness?

Let’s take a closer look at the allocation of school places as an example. This works very differently in different cities and states. The trend is toward centralized procedures within a school district. Here, families are given the option of a list of, say three, preferred schools in the school district. The schools and school types are also given priorities for students, such as by achievement level or length of commute, as well as a maximum enrollment capacity. These data are now to be used to assign students to schools as well as possible. But what exactly constitutes good is not so clear-cut. There are many possible criteria, which are often incompatible with each other and cannot be fulfilled perfectly at the same time. Accordingly, there are several approaches to the allocation of school places.

Boston Algorithm – First Choice Counts!

A simple procedure is as follows. One first tries to satisfy the initial requests of all students. In doing so, some schools will be over-subscribed and will have to turn away applicants. If applicants have equal priority with a school, then random selection is only fair. But what happens to those rejected? They are assigned to their second choice, if possible, in a second round. Some of these schools are also already full, and some applicants are rejected again. After another round, all applicants are either assigned or have worked through their list of choice schools and not found a spot.

This process, called the Boston Algorithm or Immediate Acceptance Algorithm, is or has been used in many school districts with centralized school placement around the world. Is this a good process? It depends on what the goal is. In optimization, we talk about an objective function, which must be maximized. If the objective function is the number of initial wishes fulfilled, then this procedure is optimal. That is, for each problem instance (each possible case of student preferences and schools with priorities and capacities), the maximum possible number of first preferences is satisfied. On the other hand, choice losers are accepted, that is, that there are students who do not receive any of their three choice schools, even though there would have been such an assignment.

Students and their parents know that they may not get a spot at one of their three choice schools in the Boston algorithm. Moreover, they realize that the order of schools is extremely important. It may happen that student Lilli has the highest priority at her second-ranked school A, but she is not accepted because school A has already filled its capacity in the first round – possibly with students who have a lower priority at the school than Lilli. She is legitimately frustrated. Had Lilli placed School A first on her list of choices, she would have been accepted there.

Such situations can be avoided by choosing tactically. Families are forced by the algorithm to choose the three schools very carefully. Many do not specify their school of choice, but an acceptable but possibly less popular school at #1, in order to increase the chances of securing a spot right in the first round. This results in the Boston algorithm provably assigning the maximum possible number of students to their first-choice schools, but these first choices are not the actual first choices at all!

This claim comes from Atila Abdulkadiroğlu and Tayfun Sönmez, who wrote a well-received scholarly article in 2003 citing the city of Boston as an example. That same year, the two scientists were commissioned by the city of Boston to to mathematically analyze and optimize the allocation of school places there. Boston changed the algorithm after the scientists' suggestions, and students can indicate their true school preferences in the new system without tactical concerns. Many cities around the world are successfully using this alternative process. The only loser is the city's reputation: The old algorithm was named after Boston, the city that first voted it down.

A Procedure Against Frustration

How does the new procedure differ from the Boston algorithm? Here, too, we try to assign students to their favorite, or second- or third-choice school in several rounds. The crucial difference is that here decisions are not made immediately and finally but are still adjusted over the rounds. Therefore, the algorithm is also called Deferred Acceptance Algorithm. (algorithm of the delayed acceptance decision). If Lilli could not get her first choice school A in the first round and her second choice B is already full, then she still has a chance to get a place at school B in the second round if there is a student provisionally assigned there who meets the school criteria less than Lilli.

This procedure guarantees that there will be no justified frustration; Frustration about not getting a school place that one would actually be entitled to according to school priorities. The result is also called a stable assignment. In fact, all students are assigned to the best school for them, as long as no better student has to be turned away. For Lilli, this means that she cannot be rejected by school A just because school A has already filled its capacity. The rejection in a stable assignment is only justified if school A has filled its capacity with students who all match the school priorities at least as well as Lilli.

Stable assignments and the Deferred Acceptance Algorithm were first described by David Gale and Lloyd Shapley in a 1962 scientific article. Since then, they have played an important role in numerous applications. This was also seen by the Nobel Prize Committee. In 2012, Alvin Roth and Lloyd Shapley received the Alfred Nobel Memorial Prize in Economic Sciences, colloquially known as the Nobel Prize in Economics, for the theory they developed on stable allocation procedures and their applications in the design of market mechanisms.

The Role of Mathematics

Now, is the Deferred Acceptance Algorithm actually better than the Boston Algorithm? It depends on the goals. If Lilli should be able to choose tactic-free according to her real desires, and if fairness is important to us in the sense that there should be no justified frustration, then YES.

Still, in examining the problem of school place allocation, many other questions arise that must be weighed. May there be x justifiably frustrated students if y choice losers can be avoided in return, or if z more students get their first choice as a result? How many choice losers may an additional fulfilled first choice cost?

These are decisions that are not made by us mathematicians, but by government agencies, politicians, parent representatives, etc. We provide decision support in this process, helping to weigh conflicting goals and choose appropriate procedures. We do this by analyzing problems, identifying procedures and possible weaknesses and, if possible, eliminating them by developing new procedures. Furthermore, as mathematicians, we guarantee by formal proofs that the properties of the algorithms actually apply. So that fair is really fair.

38 Failure Probabilities for Conspiracy Theories***

A mathematical model can be useful for countering the potentially harmful consequences of fake and anti-science narratives and to explore the hypothetical conditions under which a sustained conspiracy might be possible.

In conspiracy theories, individuals believe that events and power relationships are secretly manipulated by certain secret groups and organizations. Many of these supposed explanatory conjectures are unfalsifiable, unprovable, or demonstrably false, yet public acceptance remains high. Efforts to convince the public of the validity of medical and scientific knowledge can be hampered by such narratives, which create the impression of doubt or disagreement in areas where science is well established. Conversely, there are historical examples of revealed conspiracies, and it can be difficult for people to distinguish between reasonable and dubious claims.

Dr. David Robert Grimes of Oxford University, in his paper[39] developed a simple mathematical model for conspiracies with multiple actors that provides the probability of failure for any given conspiracy. The parameters for the model are estimated using literature examples of well-known scandals, and the factors that influence the success and failure of conspiracies are examined. Grimes' model is also used to estimate the probability of claims from some widely held conspiratorial beliefs; namely, that the moon landings were faked, that climate change is a hoax, that vaccinations are dangerous, and that a cure for cancer is being suppressed by interest groups.

Anti-Science Conspiracy Narratives – a Brief Overview

Conspiracy theories alleging nefarious, underhanded actions by scientists are ubiquitous. Dr. David Robert Grimes confines his work to four prominent Conspiracy Theories:

1. **NASA Moon Landing**. The successful Apollo 11 mission in 1969 put men on the moon for the first time, a landmark achievement in human history. There is a fringe group that believes the moon landings were faked for propaganda purposes. They cite alleged inconsistencies in images taken on the lunar surface.

2. **Climate Change.** Despite the overwhelming strength of evidence supporting the scientific consensus of anthropogenic global warming , there are many who reject this consensus. Many of them claim that climate change is a hoax orchestrated by scientists and environmentalists, ostensibly to generate research revenue.

3. **Vaccinations.** Conspiratorial beliefs about vaccinations are common in the anti-vaccination movement, e.g. it is believed that there is a link between autism and the MMR vaccine. This belief has reduced the uptake of important vaccinations in several countries and has led to a resurgence of diseases such as measles.

[39] D.R. Grimes, *On the Viability of Conspiratorial Beliefs*, PLoS ONE 11(1) (2016), e0147905. (+ correction)

4. **Cancer Cure**. The belief that a cure for cancer is being held back by advocacy groups is widespread. It is often used by proponents of an alternative purported cure, and the conspiracy theory claim acts as an explanatory approach to explain the complete lack of clinical evidence for such claims.

5. **COVID-19.** The coronavirus pandemic has led to a marked increase in medical disinformation and COVID conspiracy narratives on social media, such as that COVID is a hoax or deliberately manufactured, that 5G frequency radiation caused the coronavirus and that the pandemic is a ploy by big pharmaceutical companies to profit from a vaccine[40].

Derivation of Model

It is assumed in Grimes' model that, for a given conspiracy, the conspirators are generally largely concerned with concealing their activities. Further, it is assumed that an information leak (intentional or accidental) from any one conspirator is sufficient to expose the conspiracy and render it moot. This exposure of a conspiracy is a relatively rare and independent event. We can then apply Poisson statistics and express the probability of at least one leak leading to the failure of the conspiracy as

$$L(t) = 1 - e^{-\int_0^t \phi(\tau)\, d\tau}$$

where ϕ is the mean number of failures expected per unit time. This in turn is a function of the number of conspirators $N(t)$ and p, the intrinsic probability of failure per person per year. Then Φ can be given by.

$$\phi = 1 - (1-p)^{N(t)}$$

and writing $\psi = 1 - p$ for brevity, the probability of a conspiracy failure can be rewritten as a function of time

$$L\big(t, N(t)\big) = 1 - e^{-\int_0^t (1-\psi^{N(\tau)})\, d\tau}.$$

When $\phi(t)$ is a constant (homogeneous Poisson process), L simplifies to.

$$L\big(t, N(t)\big) = 1 - e^{-t(1-\psi^{N(t)})}.$$

There are several possibilities for the number of conspirators $N(t)$; the appropriate choice depends on the nature of the conspiracy.

1. If a conspiracy requires constant maintenance, then the number of conspirators required to maintain the fiction is approximately constant with time. This refers to situations where active contribution is essential to cover up an event or maintain a deception. In this case, the number of conspirators involved takes a simple form $N(t) = N_0$, where N_0 is the initial number of conspirators.

[40] D.R. Grimes, *Medical disinformation and the unviable nature of COVID-19 conspiracy theories*, PLoS ONE 16(3) (2021), e0245900.

2. If the conspiracy is instead a single event, after which no new conspirators are needed, the participants die out over time, decreasing the probability of detection. If this is the case, a *Gompertz's survival function* can be used for the function $N(t)$. If the average age of the individuals involved at the time of the event is t_e, then

$$N(t) = N_0\, e^{\frac{\alpha}{\beta}\left(1-e^{\beta(t+t_e)}\right)},$$

   where N_0 is the initial number of people involved and α and β are constants for the Gompertz curve. For humans, we can choose $\alpha = 10^{-4}$ and $\beta = 0.085$ to describe human mortality.

3. When conspirators are quickly removed due to internal disputes or otherwise, (an action that is itself arguably a meta-conspiratorial event), there may be circumstances where we can model $N(t)$ as an *exponential decay*. If members are removed rapidly and only half are left after a period t_2, then the decay constant $\lambda = \ln 2/t_2$ and the number of conspirators at a given time is

$$N(t) = N_0\, e^{\lambda t}.$$

   This equation is based on the assumption that removing confidants quickly does not change the probability of detection per confidant.

We see that an increase in $N(t)$ always leads to an increase in $L(t)$, no matter what form is chosen for the conspirator density. The time failure rate is a bit more complicated; for the constant case, L will increase monotonically with time. If non-constant forms are used instead, L will be non-linear with time. The time t_m at which L is a maximum in these cases is obtained by solving $\partial L/\partial t = 0$, which yields the following equation.

$$1 - \psi^N(t_m)\Big(1 + t_m \ln(\psi)\frac{\partial N}{\partial t}\big|_{t_m}\Big).$$

This equation can be solved by numerical methods. The maximum failure probability is then $L(t_m)$ and the The form of $N(t)$ clearly shapes the dynamics of the problem, as shown in Figure 58.

Parameter Estimation

To use the Gompertz model, the parameters must be estimated. In particular, the parameter p, the probability of an intrinsic leak or fault, is extremely important. By definition, details of the conspiracy are rarely known, but we can estimate the parameters very conservatively using data from exposed examples. To estimate, we consider exposed conspiracies where sufficient data on the duration and number of conspirators are publicly available.

As an example, let us consider the NSA (National Security Agency) PRISM affair, i.e., spying on civilian Internet users, which involved tapping fiber optic cables, intercepting phone calls, etc. With the available data, we can conservatively estimate p. In doing so, we assume that after duration t, when conspiracies are uncovered, the probability of their failure is $L \geq 0.5$. A lower bound for p is then.

$$p \geq 1 - \sqrt[N(t)]{1 - \frac{\ln 2}{t}}.$$

There is considerable and unavoidable uncertainty in some of these estimates, particularly in the number of people who have full knowledge of the event. In the case of PRISM, the 30,000 figure is the total number of NSA employees. In reality, the fraction of employees who would have knowledge of this program would likely be much lower, but we take the upper number to minimize the p estimate. Given the short time period, we further assume that the number of conspirators remained roughly constant over the period before the event was revealed. In addition, the lifespan of the conspiracy is not always clear – in the case of the NSA, the estimates are about 6 years. Further, we use the estimated values $p = 4.09 \cdot 10^{-6}$ and $\psi = 0.99999591$.

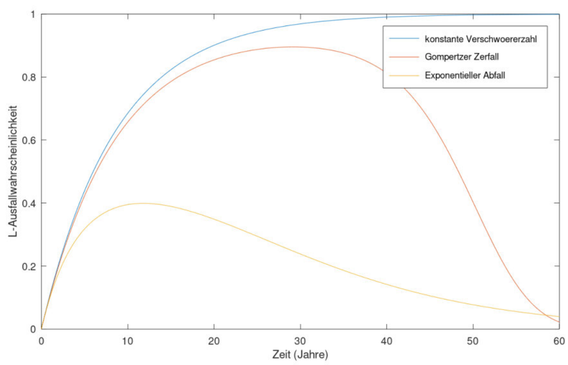


Figure 58: Projected probability of failure L for a conspiracy with 30,000 initial conspirators and $p = 4.09 \cdot 10^{-6}$ (example PRISM) and half-life $t_2 = 8$.

39 Fast Satellite Communication with π *

The number π dominates modern satellite communications.

„Superfast Internet for everyone!“. This advertising slogan is everywhere these days. But is it really possible for everyone? Internet providers don’t really like the idea of running an expensive Internet cable to a small village in the mountains of Siberia. These people have to make do with a satellite connection, and it’s not always very fast. This is because more people in remote areas of the world want to use the same connection. Moreover, the connection in Siberia also slows down if the satellite is flying over the South Pole at that very moment!

People in remote places around the world can consider themselves lucky recently, thanks to the SpaceX company and its ambitious project called Starlink. This company is in the process of building a large network of multiple satellites to provide a fast and constant connection to anywhere on Earth. To guarantee a fast connection everywhere in the world, SpaceX wants to eventually launch 42 thousand satellites into space! In the animation you can see how this will look like.

At the time of writing (February 2021), there are about 1000 satellites in orbit around the Earth. This orbit around the Earth is also called an orbit.

To avoid accidents, the launch of 42 thousand satellites must be well thought out. That’s why it’s important to be able to calculate exactly what the satellite’s orbit will look like. And that’s where the constant π comes into play! In the formula below, you can see that π is used to calculate the period (p) of an orbit:

$$p = \frac{2\pi r\sqrt{r}}{\sqrt{Gm}}$$

In other words, this formula calculates how long a satellite needs for one orbit around the earth. This formula can also be rewritten into a new formula, which can be used to determine the radius (r) of the satellite's orbit:

$$r = \left(\frac{p\sqrt{Gm}}{2\pi} \right)^{\frac{2}{3}}$$

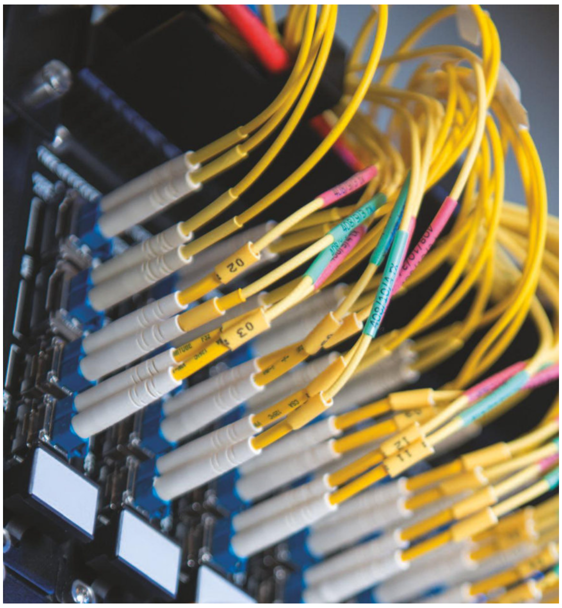
With this radius one is able to determine the altitude of the satellite. Once the period (p) and the radius (r) are calculated, it is possible to determine the velocity (v) of the satellite. Also here π plays a role:

$$v = \frac{2\pi r}{p}$$

So the constant π is used in the calculations to keep satellites cleanly in orbit around the Earth. SpaceX's Starlink network is one of the many applications, in which satellites play a role. We also owe navigation in our cars or on our phones to the technology behind satellites.

40 Traffic Nodes of Digital Data Packets*

Frankfurt am Main is home to Germany's largest Internet hub, the *German Commercial Internet Exchange* (DE-CIX). DE-CIX contributes to a fast, reliable and cost-effective Internet.



As a consumer, when you watch a movie from a digital media company like Netflix, you get that movie from your own Internet provider. But how does the Internet provider get the movie from Netflix? That happens at the Frankfurt-based Deutsche Commercial Internet Exchange, or DE-CIX (https://www.de-cix.net/). Many major cities around the world have their own Internet Exchanges. These Internet Exchanges are designed to allow connected customers to exchange digital data with each other cheaply, efficiently and reliably. Customer A does not do this directly and separately with parties B, C and D, but simultaneously with B, C and D via such a transport node.

DE-CIX is the world's largest commercial Internet node. It was officially founded in 1995 and was located in the former Telegram Post Office, which was already connected to the fiber optic network. DE-CIX is the fourth largest Internet service provider (ISP) in terms of the fourth largest Internet exchange in the world in terms of traffic volume (as of April 2019).

What the Port of Hamburg is to shipping and Frankfurt Airport is to aviation, DE-CIX is to Internet traffic.

Ethernetswitch

At DE-CIX, network engineers have to optimize the flow of data between all customers. They use devices that receive data from a sender on the front end, e.g. Netflix, and send the same data to the correct receiver on the back end, e.g. your Internet provider.

These devices are called *Ethernet switches*: big cabinets full of cables. DE-CIX has access Ethernet switches that connect directly to a customer and core Ethernet switches that connect the access switches together. The advantage of using two types of switches is that it allows DE-CIX to create multiple ways to send data from point A to point B. This reduces the risk of congestion. This reduces the risk of congestion during heavy traffic and increases reliability if one of the roads unexpectedly fails. This involves solving the problem of how to split customers across the first type of Ethernet switch in such a way that the cost is as low as possible, while the speed and reliability of the traffic must be as high as possible.

To solve this problem, one can use mathematical methods of *combinatorics*. Computing all possible connections between all connected clients would take an impractical amount of computing time. Combinatorics helps to find good solutions quickly. Yet many combinations that are possible in theory make no sense in practice. For example, it would be very expensive to connect just one large customer to an access exchange, because all of that customer's traffic would then have to be transported from DE-CIX to other exchanges. The algorithms that will be used in the software can take this into account. A simple sort that makes the problem much easier is to connect customers that exchange a lot of data with each other to the same Ethernet switch as much as possible.

Graph Theory

A second problem that is solved mathematically here has to do with proactively checking whether connections between Ethernet switches still work properly. For this purpose, *graph theory* is used, the mathematical theory of networks connecting nodes. This can

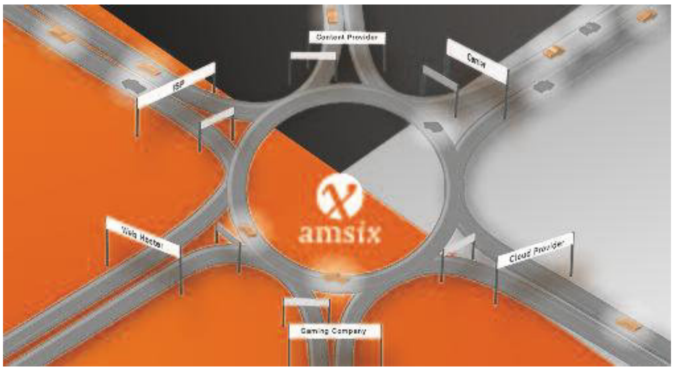


Figure 59: DE-CIX works like a traffic circle for Internet traffic: Via DE-CIX, customer A can reach all other customers simultaneously.

be used to determine how all connections can still be verified with as few test devices as possible. The algorithm used for this is a variant of the famous *Dijkstra algorithm*, which is also used, for example, in navigation software in cars.

The amount of Internet traffic handled by DE-CIX grows significantly every year, so more Ethernet switches have to be installed every year. This usually happens during the quiet summer months. In 2019, the peak capacity of DE-CIX is no less than eight terabits per second - on the order of ten thousand times faster than what you have at home as a consumer. And thanks to significant excess capacity and the distribution of Ethernet switches across countless physical locations in Frankfurt and many other exchange points in Germany and around the world a fast and robust Internet infrastructure is created. Due to lockdown measures during the COVID-19 pandemic, traffic at DE-CIX increased continued to increase since early/mid-March 2020, especially in video communications and gaming.

41 The Mathematics of Jostling*

The dynamics of pedestrian flows is a young and growing research field in which different scientific disciplines are working. Mathematics has an important integrating role in this interdisciplinary field. Applications range from the organization of large events, to the optimization of traffic facilities, to the design of escape routes for schools or museums.

Armin Seyfried has been working on the dynamics of passenger flows for 15 years. Mira Küpper is working on her doctorate on the analysis of walking paths and waiting behavior on train platforms. For this, they cooperate with scientists from engineering, mathematics, computer science and social psychology. The collaboration with mathematicians is important because it leads to new impulses in data analysis and modeling.

Although the connection between pedestrians and mathematics is not directly obvious, there are many mathematical methods that play a central role in pedestrian dynamics research and from which a variety of practical applications arise. How do you get several thousand visitors in and out of a soccer stadium? How can you calculate how wide the escape routes from a building have to be so that in the event of an evacuation as many people as possible can leave the building in a short time? Where does this cause congestion and how can this be prevented?

The difficulty with such questions is that unlike car traffic, where there are concrete traffic rules (lanes, right-before-left, etc.) pedestrians can move in a criss-cross fashion. In order to study pedestrian movements in detail, pedestrian experiments are often conducted under „laboratory conditions“, see Figure 60. Numerous such experiments have already been carried out in the context of various research projects of the „Computer Simulation for Fire Protection and Pedestrian Traffic“department of the Bergische Universität and the „Institute for Advanced Simulation“ of the Forschungszentrum Jülich. In the context of pedestrian dynamics, this means having several hundred volunteers jostle through corridors and constrictions in a large hall in order to be able to scientifically observe the movements of people. During the experiments, the volunteers wear colored caps and are filmed by cameras from a bird's eye view, so that the walking paths of the people can be traced in retrospect. Various experiment setups can be used to investigate when subjects form an orderly queue and which factors lead to jostling.

The data collected by experiments are analyzed in terms of density, i.e. how many people were present in a square meter area, the speed of the subjects, and the flow, which describes the number of people who crossed a given line, e.g. a bottleneck, in a fixed time interval. By comparing the results of different experimental setups, it is possible to analyze the effect of changing the setup (e.g. corridors with different widths) or the instruction of the subjects (e.g. „Only the fastest get a good seat“).

Since experiments can only be carried out on a manageable scale and thus for simplified questions, and it would, for example, be far too time-consuming to test in an

Figure 60: An experiment with 700 test persons in which congestion at intersections was investigated. The experiments took place in 2013 as part of a research project on major events in the exhibition halls in Düsseldorf. (Source: Research Center Jülich / Photo: Marc Strunz-Michels)

experiment how long it takes to evacuate an entire soccer stadium, such complex questions are investigated using simulations. A simulation is a computer model that attempts to reproduce reality.

Mathematical and physical computer models are needed to simulate and thus predict the movement of pedestrians. Mathematics can be used to find ways to translate the behavior of people into equations and models, see Figure 61. For example, the model must „learn“which factors cause previously nondrushing individuals to become jostlers. To calibrate and validate these mathematical models, reproducible results from laboratory experiments are used. This ensures that reliable predictions of where jams may occur can be calculated. After all, any computer model can only be as good as the data used to calibrate it.

With good computer models for the movement of people flows, the safety of people in everyday situations can be increased. Therefore, this is a research area with relevance to everyday life to which every individual can contribute his or her own experience.

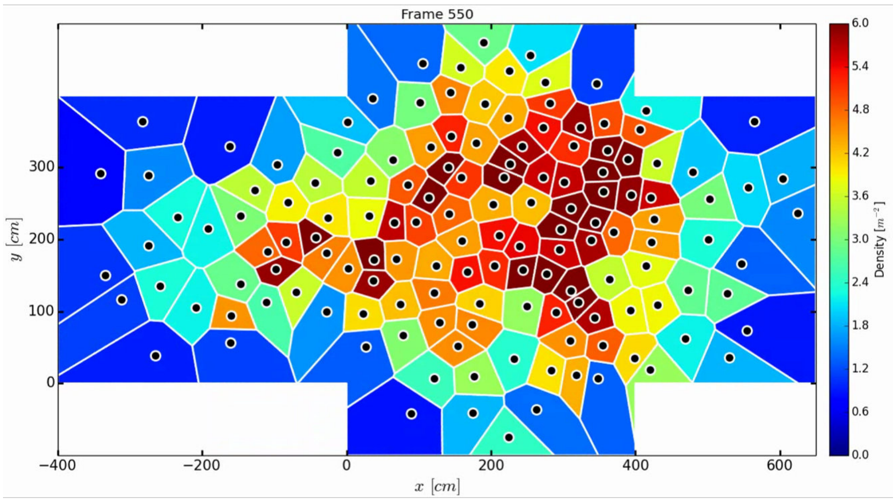


Figure 61: Voronoi decomposition of the space for the determination of individual densities. The raw data, black dots, are positions of pedestrians. They were measured in an experiment similar to the one shown in Figure 60. The values show that the density can be more than 6 people in one square meter.

42 With Mathematical Optimization to Safe and Climate-Friendly Flights***

An exciting challenge for the mathematical theory, research, and practice of optimal control applied to a problem of high societal, environmental, and economic relevance.

A report from the scientific practice of Prof. Dr. Ekaterina Kostina (Institute of Applied Mathematics and Interdisciplinary Center for Scientific Computing of the Ruprecht-Karls-University Heidelberg).

When you look up at the sky and see strange patterns of condensation trails (see Figure 62), have you ever wondered how many flights there are like this every day?

Figure 62: A quiet afternoon in Heidelberg.

Let's look at the figures published in the annual report[41] of EUROCONTROL, the organization responsible for the central coordination of air traffic control in Europe: It's hard to believe, but in 2019, before the COVID-19 pandemic, the number of flights in Europe reached 11.1 million. In Germany alone, more than 9,000 planes took off from, landed in or flew over Germany on any given day in 2019. Then, in the following year, there were fewer than 5 million flights due to the pandemic in Europe, but experts believe that if the vaccination campaign is successful, air traffic could return to 2019 levels as early as 2024. In the long term, further strong growth is even expected. Even in times of pandemic, there is no denying that air travel has become a matter of course in our society, making people more mobile than ever. However, growing air traffic brings with it many challenges, not least for air traffic management.

The primary goal of air traffic management is flight safety. This means that air traffic controllers must ensure that all aircraft keep sufficient distance from each other at all times during each flight. But that's not the only difficulty. Air traffic management must

[41] https://www.eurocontrol.int/publication/performance-review-report-prr-2019

also deal with delays in the system. According to EUROCONTROL's 2019 annual report, one-tenth of all flights in Europe were already delayed. This may not sound like much, but it causes significant disruptions to overall flight schedules, complicating flight safety and causing congestion in the air and at airports. Delays are not only inconvenient for passengers, they also lead to longer flight distances, a reduction in operational efficiency, an increase in fuel consumption and thus an increase in CO_2 emissions.

And that brings us to the environmental issues associated with air travel and the EU's ambitious climate targets. You may know that to combat climate change and global warming caused by CO_2 emissions, the European Union has agreed to strict sustainability targets for climate-neutral and zero-emission mobility by 2050. According to the „European Aviation Environmental Report 2019"[42], aviation currently accounts for 3 % of global greenhouse gas emissions, but long-term projections see the relative share of aviation continuing to rise as it continues to rely on fossil fuels in the medium term.

As an example, a typical aircraft with two jet engines consumes 2,700 kg of kerosene during a one-hour flight with 150 passengers, resulting in emissions of about 8.5 tons of carbon dioxide. For flights in Europe, for example, this then adds up to a total of 171 million tons of CO_2 in 2016, almost twice as much as in 1990.

Emission reduction strategies include replacing the existing aircraft fleet with lighter and more fuel-efficient aircraft, developing more efficient jet engines, carbon-neutral biofuels, hydrogen or electric engines, or reducing drag. But efficient air traffic management systems and planning more fuel-efficient flight routes can also play an important role in reducing emissions, since air traffic management can affect about 6 % of total „gate-to-gate fuel consumption".

Especially during the high-consumption takeoff and landing phases, air traffic controllers and flight planners can play an important role in reducing the load. To overcome all these challenges, air traffic controllers need high-quality decision support tools to efficiently manage the upcoming traffic flow. The development of such a tool is hardly possible without modern mathematical methods of optimal control.

Historical Excursion

Modern methods of optimal control have their roots in the theory of calculus of variations, which began with the so-called „brachistochron" problem (Greek: brachistos=shortest, chronos=time). The problem was formulated by Johann Bernoulli in 1696: What is the curve of the shortest falling time of a point of mass rolling frictionlessly from a point A to a point B under the influence of gravity? The solution of the problem (which, by the way, is neither the direct line connecting points A and B, nor an arc of a circle) by Johann Bernoulli and his brother Jacob, and later by such eminent scientists as Gottfried Wilhelm Leibniz and Isaac Newton, Leonhard Euler and Joseph-Louis Lagrange led to the works of Adrien-Marie Legendre and Karl Weierstrass, and finally to modern optimization theory. The need to solve real applied problems, e.g. of space travel, led to

[42] https://www.eurocontrol.int/publication/european-aviation-environmental-report-2019

the emergence of Pontryagin's so-called maximum principle in optimal control theory in the early 1950s. The maximum principle describes the conditions that must be satisfied by the optimal solution and has proved extremely valuable in applications.

As early as the 1980s, these methods were used to optimize nonlinear processes, sometimes with impressive results. However, optimal control must be determined „indirectly", from a very difficult boundary value problem with adjoint equations to set up and solve. As the practical applications became more complex, this approach proved to be inadequate to solve the problems. Therefore, it was replaced by methods in which optimal controls „are computed directly" as decision variables. Modern methods are based on the „direct multi-objective method", a method introduced in the 1980s by the Heidelberg mathematician Hans Georg Bock. At the Interdisciplinary Center for Scientific Computing at Heidelberg University, Ekaterina Kostina and other research groups are developing and improving highly efficient methods of this type for many different problem classes and complexities.

How can we Apply Optimal Control Methods to Air Traffic Management Problems?

We discuss an optimal control problem for determining consumption-optimal and collision-free trajectories for multiple aircraft in cruise and approach. Thus, we study an airspace sector S in the air during a given time period and aim to, find the optimal way to navigate all aircraft through S while maintaining safety margins. First, we begin by mathematically modeling the movements of all aircraft in the given sector. In our application, the model consists of a nonlinear system of differential equations of flight dynamics that describes how the „states", i.e., flight direction, velocity, and spatial position evolve in time.

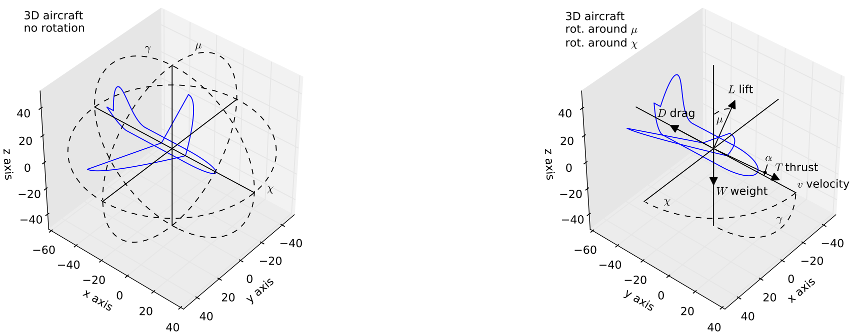


Figure 63: Representation of the conditions and forces acting on the aircraft.

The three-dimensional trajectory can be described by seven states and three controls and represents the translational dynamics of the aircraft. The seven states in Figure 63 are the (x, y) position, the altitude z, the velocity v, the heading angle χ, the climb angle γ, and the aircraft mass m. The three controls are the angle of attack α, the bank angle μ, and the thrust lever position δ. The angle of attack α is the angle between the wings and the incoming air. The angle of attack is controlled by the flap on the wings and has the function of balancing the aerodynamic forces and moments around the center of gravity of the aircraft. The bank angle μ is effected by the ailerons and has the task of changing the direction of the lift vector with respect to the wind speed vector at a given angle of attack. In our model, thrust is controlled by the thrust lever position δ. The thrust lever position has a value between the zero and maximum thrust generated by the thrusters.

As a result we get the following differential equations

$$
\begin{aligned}
\dot{x}(t) &= v(t)\cos\chi(t)\cos\gamma(t),\\
\dot{y}(t) &= v(t)\sin\chi(t)\cos\gamma(t),\\
\dot{z}(t) &= v(t)\sin\gamma(t),\\
\dot{v}(t) &= \frac{1}{m(t)}\Big(T_{\mathsf{max}}\delta(t) - D\big(v(t),\alpha(t)\big)\Big) - g\sin\gamma(t),\\
\dot{\chi}(t) &= \frac{1}{m(t)v(t)\gamma(t)}L\big(v(t),\alpha(t)\big)\sin\mu(t),\\
\dot{\gamma}(t) &= \frac{1}{m(t)v(t)}L\big(v(t),\alpha(t)\big)\cos\mu(t) - g\,m(t)\sin\gamma(t),\\
\dot{m}(t) &= -f_r T_{\mathsf{max}}\delta(t).
\end{aligned}
$$

The constants T_{max} - the maximum thrust, and f_r - the fuel burn rate, depend on the aircraft model. For the aerodynamic lift and drag forces, there are special models (equations) that depend on the aircraft model, atmospheric conditions, effective wing area, and air density.

The trajectories of each aircraft must satisfy several constraints, such as initial and final conditions. Other inequality constraints define the performance limits of the aircraft, such as. the maximum speed, but also ensure a comfortable flight. Many other factors are important for modeling the optimization problem. The required total time should be respected. We also want the aircraft to avoid unnecessary turns as much as possible. Similarly, we want the flight altitude to remain within prescribed limits. Furthermore, we want to keep the fuel consumption as low as possible.

Thus, the optimal control problem for each aircraft can be formulated as follows: Minimization of the objective function taking into account the equations of aircraft dynamics and all constraints.

To solve conflicts among multiple aircraft in sector S, we merge the differential equations, objective functions, and constraints of all aircraft into an overall optimal control problem. For the overall problem, we introduce the pairwise spatial constraints that ensure that the safety distance between all aircraft is maintained at all times. These pairwise constraints can be thought of as safety ellipsoids around each aircraft. Also, the threading of multiple aircraft on approach leads to further spatial and temporal distance constraints.

The overall problem is very difficult to solve due to its high dimension and the spatial and temporal constraints that have to be met, but it can be solved very efficiently with optimal control methods developed at Heidelberg University that exploit the special structures of the underlying problems.[43] [44] [45].

The solutions obtained with the computer have a very high potential for the calculation of collision-free and CO_2-efficient flight trajectories, see Figures 64 and 65

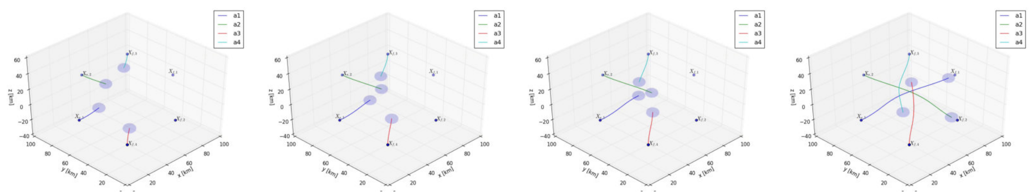


Figure 64: Collision-free and CO_2-efficient trajectories in a sector S.

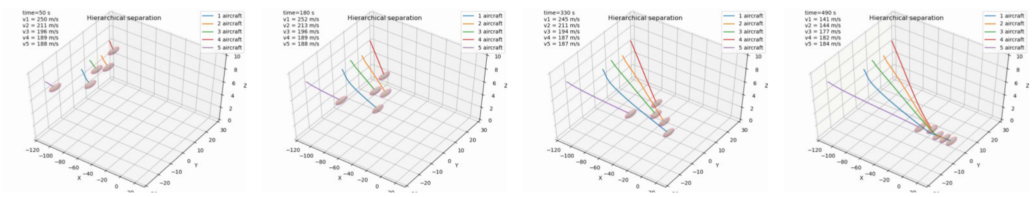


Figure 65: Collision-free and CO_2-efficient trajectories during landing approach.

Optimal solutions reduce delays in the system and their domino effects. Flight detours are reduced, enabling better utilization of airspace and further fuel savings.

Current and Future Work

Unfortunately, the model described above is still too simple for practical use and needs to be significantly expanded. Different aircraft types have specific consumption values and often differ significantly in their aerodynamic flight and control behavior. In addition, external influences that affect flight behavior, such as wind and weather conditions, must be modeled and taken into account. Such external influences are disturbances and potential sources of error that cannot be predicted accurately or at all. Wind and weather conditions can change rapidly locally and strongly influence flight dynamics

[43] M. Schubert, *Optimal Control of 3D Separation Management*, Master thesis, Ruprecht-Karls-Universität Heidelberg, 2017

[44] V. Semenov, *Multi-Stage Strategy to Numerical Trajectory Optimization for Multiple Aircraft Approaching an Airport*, Master thesis, Ruprecht-Karls-University of Heidelberg, 2019

[45] A. Raptakis, *Optimization of Flight Trajectories of an Aircraft with Respect to Fuel Consumption*, Master thesis, Ruprecht-Karls-Universität Heidelberg, 2020.

(turbulence, crosswinds, wind shear). To take these factors into account, the current situation must be evaluated and quantified very quickly and reliably, and the optimal solution must be adapted to the disturbance using real-time methods („Feedback"). We have already made tremendous progress in the development of such methods in recent years.

Our research goal is to develop, in cooperation with air traffic control institutions and airlines, high-performance decision-making systems that will help meet the diverse challenges of air traffic in the future in a mathematically optimal way. For this, we need highly motivated young scientists who are enthusiastic about mathematical optimization theory as well as about modeling the physitechnical processes of exciting applications!

43 Fewer Plane Crashes due to Better Probability Models*

Most aviation accidents are caused by a combination of engineering, human, and environmental factors. Better mathematical probability models reduce the likelihood of an accident.

Over the past fifty years, the risk of a fatal aviation accident in civil aviation has dropped spectacularly, from about 30 fatal accidents in one million flights in 1959 to less than 0.5 in one million flights in 2012. Of course, because every accident is still one too many, aircraft manufacturers and aviation authorities are constantly looking for ways to further improve aviation safety.

Safety researcher Alfred Roelen of the National Aerospace Laboratory (NLR) in Amsterdam uses mathematics to help model the probability of an aircraft crash. „We look for factors that play the biggest role in causing accidents and then reduce them," Roelen says. „On the one hand, we do this by collecting large amounts of flight data and analyzing it statistically. And on the other hand, by creating better models of the probability that something will go wrong during a flight."

There may be a technical problem, for example with a wing, an engine, or the autopilot. But something can also go wrong on a human level, for example with the pilot or the air traffic controller. In addition, weather-related problems can occur. Almost always, an aviation accident is due to a chain reaction of factors.

Ice on the Wing

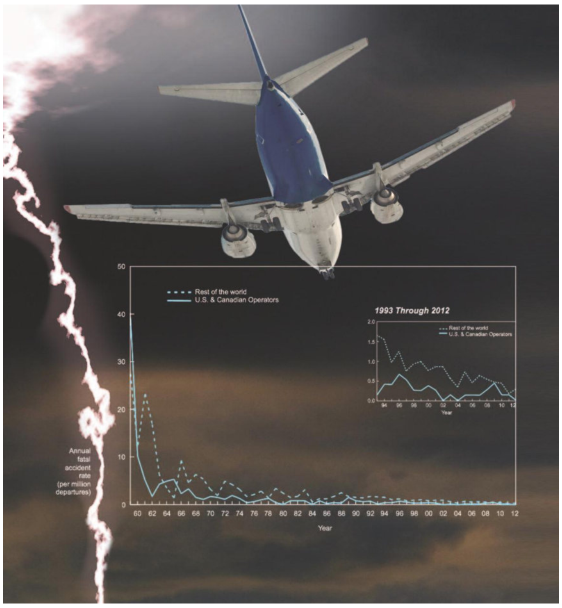
Roelen gives an example of a chain reaction that he can model mathematically: „Suppose it is snowing at an airport. According to the usual procedure, snow and ice are removed from an aircraft before takeoff. However, there is always a chance that this will not be done properly for whatever reason. What are the chances that an aircraft with snow or ice on one of its wings will take off? Then, what are the chances that – once in the air – the flow around that wing will break loose and the plane will abruptly descend?

Finally, what is the chance that the pilot will notice this in time and take appropriate action?“

European and U.S. aviation authorities impose stringent requirements on an aircraft’s probability of failure. Among other things, aircraft manufacturers must demonstrate through mathematical analysis that the probability of a catastrophic accident is less than 1 in a billion flight hours (in other words, about 1 accident in more than a hundred thousand years). Roelen: „This probability is so low that it is not possible to test the aircraft in practice. So we have to come up with something else. We do this by breaking the aircraft down into individual components and measuring or estimating the probability of failure for each component.“

Failure frequencies for the engineered components are the easiest to determine, usually through testing in the laboratory. Determining failure frequencies for human behavior is much more difficult. If an air traffic controller says „thousand feet low to the pilot during a landing procedure and he meant „two thousand feet low, that can be fatal. Psychological tests are used to determine how often people make mistakes depending on how hectic their minds are.

Roelen: „The simplest model determines the probability of human error in three situations: a person is very busy, moderately busy, or calm. It is logical that a person makes the most mistakes when he is very busy. But when he is very quiet, his attention is not as focused and he also makes more mistakes than when he is averagely busy.“



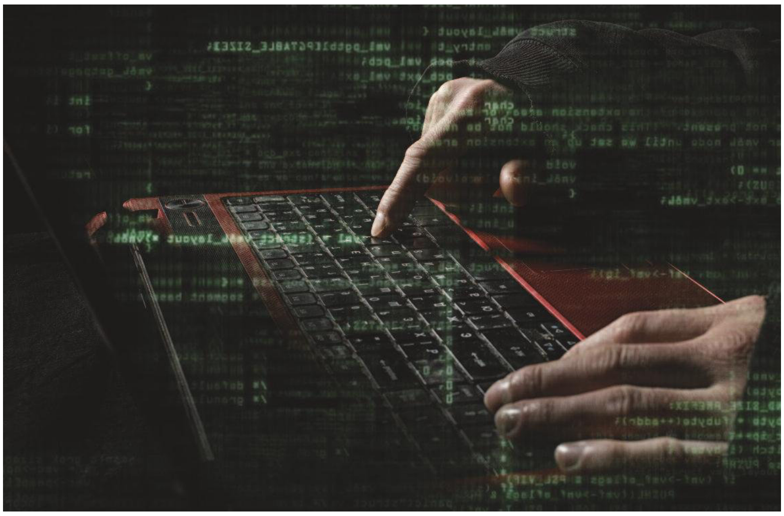
Killer

Twenty years ago, one of the biggest „killers“ in civil aviation was a type of accident called a controlled flight into terrain: An aircraft is normally en route to an airport. There is nothing wrong with the plane, but unexpectedly it flies into a mountain not far from the airport. Thanks to a detailed statistical analysis of this type of accident, Roelen and his colleagues were able to show what was the main cause.

Roelen: „At airports that did not use radio signals to guide an aircraft during landing, the likelihood of such an accident was five times higher. It is partly thanks to our work that more airports have acquired the necessary radio equipment. Nowadays, this equipment is increasingly being replaced by GPS units. Again, mathematical models are used to estimate the probability of failure."

44 Cracking Digital Protections to Make them More Secure*

These days, cryptography isn't about secrets like it was a century ago; it's mostly about math. And that modern cryptography is securing more and more everyday applications.



In 2008, researchers at Radboud University Nijmegen cracked the OV smart card. They showed how they could use it to travel for free. But the problem was much bigger than the OV chip card alone. The chip in this Dutch public transport travel card, the Mifare Classic, has been processed into more than a billion cards worldwide in more than a decade. And that includes access cards to government buildings and military installations.

It is good practice for researchers to notify the manufacturer after a security breach is discovered and give the manufacturer time to repair the product: six months to adjust the hardware and six weeks for the software. In 2008, the Radboud scientists warned the Dutch government, the internal security service and the manufacturer of the OV smart card, the Dutch company NXP. The panic was widespread. NXP tried to prevent a scientific publication by the researchers about the security breach, but the judge ruled that the truth was of public interest and allowed the publication.

Secret Key

Roel Verdult, one of the hackers at the time and now a PhD student at Radboud University, sees it as the social responsibility of digital security researchers to critically examine security. „It's very difficult to make the theoretical case that digital security is so and so strong," Verdult says. „That's why, in practice, a pragmatic approach is taken. That means that part of building a good cryptosystem is also trying to crack it. In this way, scientists can choose the best among all types of proposed protections."

Both creating and breaking cryptographic protections are based on mathematics. At its heart is a cryptographic algorithm: a computational recipe that generates secret digital keys. A commonly used method is to perform a three-number mathematical operation: first, a known number, such as a card's identification number; then, a secret value that serves as the key; and finally, a random number generated on the spot. The more difficult the key and the more random the number, the harder it is to crack and the better the security.

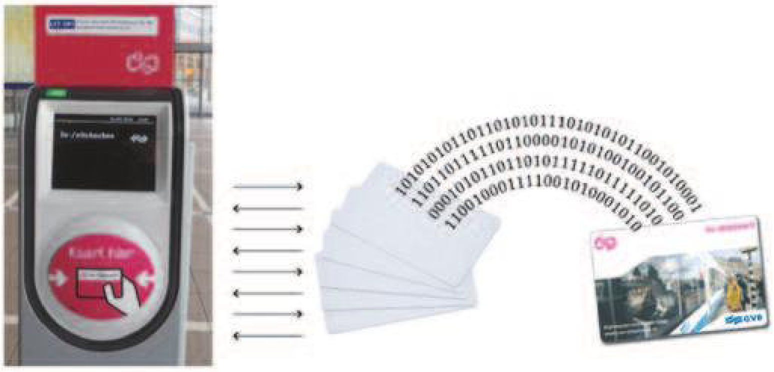
In the OV smart card example, it works as follows. As a smart card approaches a reader, it sends a unique identification number to the reader. Using this number, the reader generates a series of cryptographic keys. The card and reader use electromagnetic signals to quickly check whether they know the secret key. If they do, the traveler is checked in or out.

„The algorithm used in the Mifare Classic, and thus in the OV smart card, was already uncertain at the design stage", explains Verdult. „But because the algorithm was kept secret, it took us some time to discover the design flaws. Immediately after that, we warned everyone about the existing weaknesses. Although we were the first to openly talk about the problems, it is not unlikely, that these weaknesses were already being secretly exploited by others.

In fact, the starting point should be that the security of a cryptosystem should only benefit from the key, not the algorithm, Verdult says. „The random and secret numbers are always recalculated, while the algorithm remains the same. A good algorithm does not become insecure if it is public. In fact, anyone can then verify that it is a good algorithm. And anyone can suggest improvements to make it even stronger."

New Passport

The researchers' recommendations for a greatly improved cryptosystem were immediately heeded by the Dutch government. The government is now in the process of introducing a secure national passport for access to ministries and other important buildings. A good example of the usefulness of their work, Verdult said. But the OV smart card is a very different story. „Actually, there are few places to improve security there. But at the core, there's still the same insecure Mifare Classic algorithm. Really strange, because even the manufacturer of the Mifare Classic advises against buying the chip."



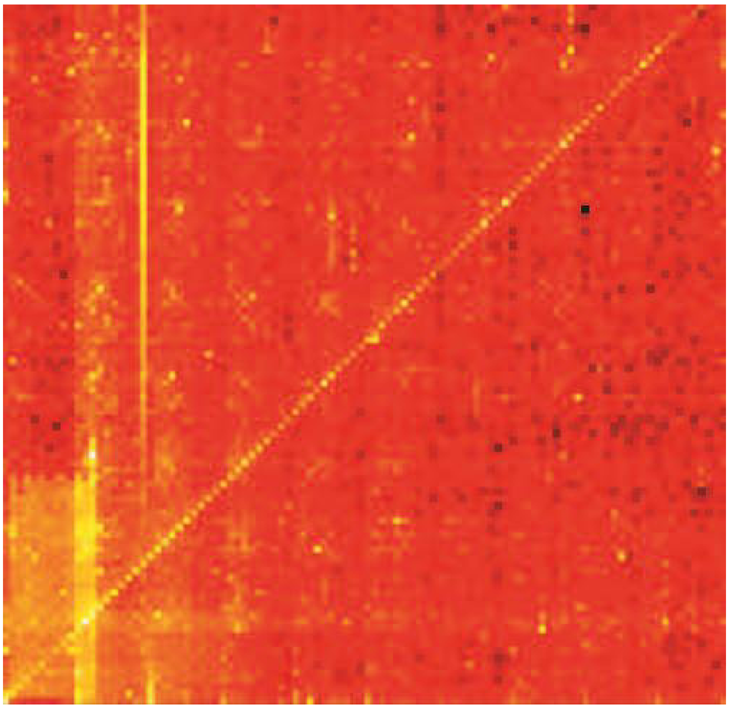
After the hacking of the OV smart card in 2008, you heard some people say as an excuse that any system can be hacked. But Verdult stresses that you will never hear that statement from a crypto researcher: „There are many algorithms that are secure. A well-known example is the *Advanced Encryption Standard* (AES), which was introduced in 1998 and after more than fifteen years is still uncrackable.“ It is the job of crypto researchers to figure out which algorithms are secure and in which applications they can best be used.

At the University of Wuppertal there are learning stations on cryptography, the Spion-Camps, cf. https://ddi.uni-wuppertal.de/www-madin/material/spioncamp.html.

45 Patterns in PIN Codes*

What do you do when you get to choose the PIN code for your bank card? Or if you can specify a four-digit code for the locker in a hotel room? You're probably more predictable than you think.

Data analyst Nick Berry collected millions of pieces of data from leaked passwords and PIN code databases. Eventually, he had a collection of millions of users who chose a four-digit code. In total, ten thousand possible combinations occurred at least once, but some were much more common than others: for example, one in ten people chose 1234 as their super-secret code. A thief enters one of five safes by trying only the five codes 1234, 1111, 0000, 1212, and 7777.



Discover the patterns, and choose a better code from now on!

Berry made this fascinating picture showing how often the combinations occur. On the horizontal axis are the first two digits of the PIN code from 00 to 99. On the vertical axis are the last two digits of the PIN code, again from 00 to 99. Each square therefore represents exactly one PIN code. The more often the PIN code is dialed, the lighter it is. So 1234 is white and the least common PIN 8068 is black. This picture shows all sorts of interesting things about how people choose their PIN codes.

46 Benford's Law – How Mathematics Exposes Number Fakers**

A simple mathematical law helps detect counterfeiting in various fields.

The history of Benford's law begins with a 1938 article by Frank Albert Benford[46] and shows how even theoretical results can have concrete practical applications, such as detecting tax fraud.

Frank Benford was interested in patterns in number data and observed a digit distortion in the distribution of the first or leading digits and wanted to explain this mathematically. Especially in an age where we are constantly bombarded with huge amounts of data, this discovery is very important. This regularity is also called „Newcomb-Benford Law" (NBG), since it was actually already in 1881 by the mathematician Simon Newcomb[47] after he had noticed that the pages used in logarithmic tables those pages with the one as the first digit were dirtier than other pages, that is, they had been used more frequently. And so Benford wrote in his article:

> *There may be, in the relative cleanliness of the pages of a logarithm table, data on how we think and how we react when dealing with things that can be described by means of numbers.*
>
> Frank Benford, 1938

Before we can describe the law, we must first establish the notation. Every positive number x can be written in the decimal system as $M(x) \cdot 10^E$, where $M(x) \in [1, 10)$ is the so-called mantissa and E is the exponent. The integer part of the mantissa (the number before the decimal point) is called the leading digit or first digit. The number 14568.79 is written in scientific notation as $1.456879 \cdot 10^4$. The mantissa is 1.456879, the exponent is 4, and the leading digit is 1.

The most obvious assumption would be to claim that, given a general data set, all numbers are equally likely to be the leading digit. So we would then assume that we should observe a leading digit of 1,2,...,9 in about 11 % of the cases. Here we have assumed that each number occurs 1/9 of the time and not 1/10 of the time, since 0 is only the leading digit for the number 0). The content of Benford's law is that this is often not the case; in particular, in many situations we expect the *leading digit* d to occur with probability about $P(d) = \log_{10}\left(\frac{d+1}{d}\right)$ bzw. $\log_{10}(d+1) - \log_{10}(d)$ which means that the probability of a first digit of 1 is about $\log_{10} 2 \approx 30,1\,\%$, while a first digit of 9 occurs about $\log_{10}\left(\frac{10}{9}\right) \approx 4,58\,\%$ of the time.

[46] Frank Benford, *The law of anomalous numbers*, Proceedings of the American Philosophical Society, 78, No. 4 (1938), 551-572.

[47] Simon Newcomb, *Note on the frequency of use of the different digits in natural numbers*, American Journal of Mathematics, 4 (1881), 39-40,

A set of numbers satisfies Benford's law for the leading digit if the probability of observing a first digit of d is approximately is equal to $P(d) = \log_{10}\left(\frac{d+1}{d}\right)$. Finally, instead of examining only the leading digit, we could examine the entire mantissa. So instead of asking for the probability of a first digit of 1 or 2 or 3, we now ask for the probability of observing a mantissa, for example, between 1 and 3. This generalization is often referred to as Strong Benford's Law: We say that a data set satisfies Strong Benford's Law if the probability of observing a mantissa in $[1, m)$ is $\log_{10} m$. Note that Strong Benford's Law implies Benford's Law: the probability of a first digit of d is simply the probability that the mantissa is in $[d, d+1)$, also $\log_{10}(d+1) - \log_{10}(d)$.

Benford's law can be used in a variety of ways, such as by accountants in detecting tax fraud, falsification of financial statements, or generally in detecting irregularities in accounting. It can also be used in science to detect data falsification. It has also been used in investigations of significant irregularities in elections, e.g. in the 2009 presidential elections in Iran.

47 Machine Learning and Mathematics in Autonomous Driving**

Neural networks and mathematics improve autonomous driving.

Machine learning, in particular deep learning methods based on so-called deep neural networks, are used in autonomous driving and its precursors mainly in three application fields:

1. In environment perception using camera and sensor data processing.
2. For learning and implementing a driving strategy.
3. For monitoring the interior

In these application fields, the new deep learning algorithms, developed in the 2010s, have developed prediction accuracies that seemed unattainable with previous methods. The 3 application fields of machine learning are described in more detail below:

Environment Perception through Sensor Data Processing

Three main sensors play a major role in environment perception: camera, LiDAR and radar sensors. More exotic variants such as heat sensors are currently less common. Based on the sensors, other road users, the road, obstacles, etc. are detected. The totality of the detections is transferred into a 3D environment model and provides the basis for subsequent journey planning and driving strategy.

Camera. Camera data is represented by 3 color values in each pixel, resulting in an HD image containing several million numerical values. Using these pixel values, we would like to detect objects in the image and assign them to a class (car, truck, bus, human, bicycle, etc.). Mathematically, we want to find a function that assigns positions and classes of all objects in the image to the given pixel values. A (classical) manual modeling based on self-devised rules does not achieve the desired detection accuracy here. Deep learning methods have revolutionized this application.

Deep Learning figuratively uses templates of hierarchically organized functions with a large number of free (learnable) parameters. The parameters are adapted in a mathematical optimization to learn a given functional context (more information is later in this section). For example, a deep neural network can learn to assign each image pixel its class membership. This is illustrated in the figure 66. Here, a color is assigned to each class for illustration. To learn such a function, large amounts of so-called labels are needed. Currently, neural networks are trained for this task with orders of magnitude of 10,000s of images, for each of which humans have outlined the objects in the image with edge features. The acquisition of such labels is time-consuming and costly, but currently still unavoidable for accurate environment perception.

Figure 66: Example of a so-called semantic segmentation, calculated by a deep neural network. For each image pixel a class membership is estimated, by an assignment of classes to colors such a segmentation image is created.

Radar. Another sensor widely used in the automotive industry is radar (**RA**dio **D**detection **A**nd **R**anging). Although radar does not provide images in the classical sense, the sensor has several advantages: the quality of the measured data does not depend on visibility, and so radar provides usable data even in scenarios where camera often has problems, such as under poor lighting, fog or in bad weather. Another advantage is that radar can directly measure not only the position but also the speed of the object. Radar works on the basis of electromagnetic waves: a signal is emitted and when this hits an obstacle it is reflected back. This echo is then compared with the transmitted signal to measure the time offset between the two signals. The time offset provides information about the distance of the reflecting obstacle.

Several measurements even allow to measure the distance change over time, i.e. speed. By using several radar antennas side by side, it is additionally possible to extract information about the observation angle and height of a reflecting object. Suitable mathematical procedures convert the received time singals into the position data of the reflecting targets, so-called detection points, which form a basis for recognition of the objects (pedestrians, cars). Over the years, the processing of radar data was dominated by classical signal processing methods, but in recent years, machine learning techniques have also prevailed in this area. However, since radar data is very different from images captured by cameras, Deep Learning technology must also be adapted for this purpose – a challenge that Aptiv's Wuppertal office is also working on. Understanding the mathematics behind these technologies is essential for developing suitable Deep Learning as well as classical signal processing methods.

Lidar. In addition to the camera and radar, a lidar (**Li**ght **D**etection **A**nd **R**anging) sensor is also of great importance in autonomous driving. Similar to radar, lidar also measures back-reflected signals: but in this case, instead of an electromagnetic wave, it involves a beam of light that is reflected with a corresponding delay. This time delay is also converted into the location information of the reflection point. By several emitted light beams a multiplicity of reflection points can be generated, which form a three-dimensional point cloud, in which one (depending upon quality of the lidars) can recognize the environment well. The multitude of reflection points and the accuracy of position measurement is one of the clear advantages of lidar sensors. On the other hand, the distribution of point data in three-dimensional space represents a challenge for many of the deep learning methods, which were often developed specifically for two-dimensional images. When developing new algorithms, the running time of the algorithms must be considered as well as the quality of the detection: if the algorithm takes too long, the results would be available too late to react appropriately in the situation.

Whether camera, radar, or lidar, each of these sensors brings with it some disadvantages in addition to advantages. This is what makes another area of research important: sensor fusion. With several different sensors, the risk of all sensors failing at the same moment can be minimized.

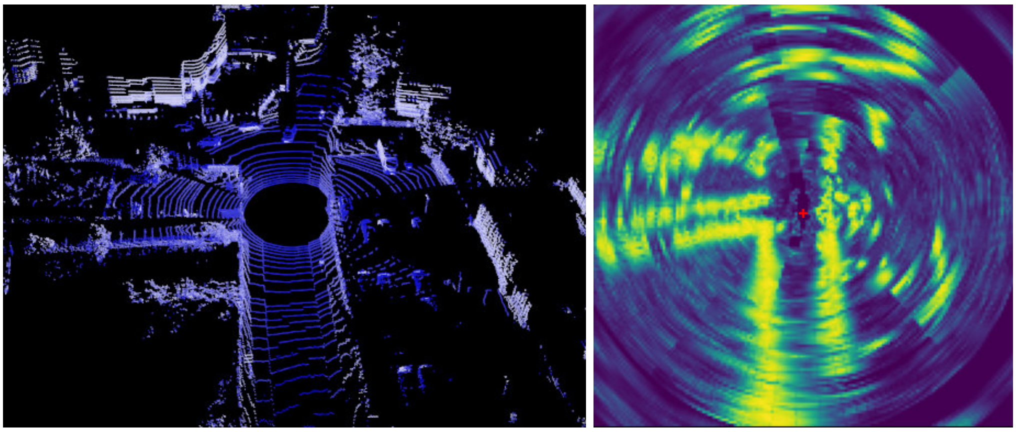


Figure 67: Example of data from lidar (left) and data from radar (right).

Driver Assistance and Automated Vehicle Control

Using information extracted from sensor data about other road users (such as pedestrians, bicycles, cars) and about the environment, the vehicle can take automated action. Depending on the level of automation, this can involve e.g. a warning to the driver, automatic emergency braking or even full control of the vehicle. Especially for more complex tasks approaching full automation, it can be difficult to program the car's behavior using only rules, and another part of machine learning comes into play: reinforcement learning. This method involves making optimal decisions based on all the information available at the time (environment, positions and directions of movement of other road users, but also all such available data from the past). Not only the immediate, but also the further consequences of this decision have to be taken into account.

To represent the decision as a mathematical function of all available information, one again often accesses the deep networks. The consequences of this decision are then evaluated by a positive numerical value if they are desirable (such as arriving at the destination), and by a negative value if they are undesirable (such as causing an accident). With regard to safety, this network is usually trained only in a simulation. During training, the network is confronted with a traffic situation and makes a recommendation based on the available information. The recommended action can then be executed in the simulation, but it does not have to be: with certain probability, the program makes a different decision in order to investigate the unknown consequences. Based on the consequences, the parameters of the neural network are then changed so that in the future, in similar situations, the network provides recommendations with the most positive-valued consequences possible.

Another method of learning optimal control is called imitation learning. In this method, recorded (non-autonomous) trips are used as data: during training, the neural network then receives the recorded data from a scenario as input and tries to imitate the recorded decision of the human driver based on this data.

Monitoring the Interior

Similar to the detection of the environment using cameras, the detection of people and objects in the interior is also of great importance, both for automated driving and for assistance systems. Cameras and deep learning are also used in the interior, see Figure 68. For example, the detection of people, the recognition of their condition, the detection of child seats, children and pets are of higher interest. In preliminary stages of automated driving, for example, it is necessary to determine whether the passenger in the driver's seat is in condition to take the wheel. Acute symptoms of illness could also be detected. The vehicle can also indicate if a child or pet is left in the car. In mobility solutions that are expected to complement public transportation with smaller autonomous group cabs in the future, it also plays a role in monitoring passengers to detect assaults, for example.

Learning Functional Relationships using Deep Learning

Now that we have seen many examples of the application of Deep Learning in automated driving, let's shed some light on the role of mathematics in Deep Learning, as it goes beyond the mere approximation of functional relationships. In a neural network, the functional template of Deep Learning, mainly arithmetic operations (additions and multiplications take place) take place. Neurons form the smallest units of neural networks and receive as input a signal in the form of a number of other neurons to which there is a connection in the form of a so-called edge (as an arrow in the Figure 69). On each edge lives a so-called weight, a free parameter of the neural network, which is

Figure 68: Example of instance segmentation in the interior of a car. First, objects are localized and outlined by rectangles. Then the instance is segmented within the rectangle.

learned. The weight is multiplied by the signal (i.e., weighted) and all weighted signals are summed in the neuron and then a threshold function is applied to the weighted sum. If a certain total is exceeded, the neuron is activated and transmits a signal to the other neurons connected to it. This modeling is biologically inspired. Neural networks for the described applications are mostly so-called feedforward networks. In these, neurons are organized in groups of layers and one layer forwards signals only to subsequent layers (i.e., forward) (cf. Figure 69).
Apart from the fact that neural networks are nothing more than mathematics, various classical disciplines of mathematics play a weighty role in the following problems; here we give the most prominent examples:

1. How many learnable weights are needed for a neural network to theoretically approximate a given functional relation to a desired accuracy? The rough direction is that more weights correspond to more capacity and allow complex relationships to be represented. However, to answer this question more precisely and also quantitatively (what number of neuronals is needed?), more and more precise theory is being developed. However, as new architectures of neural networks are developed, new theory is also needed. If the neural network now has capacity in the form of free weights, how do you find the right weights? Finding optimal weights is a problem that cannot be solved in anywhere near acceptable computation time, even with the largest computers in the world. Efficient algorithms (called heuristics) that provide the best possible solutions to the problem of finding weights are the subject of current research.

2. How much data does it take to approximate a given functional relationship to a desired accuracy? Roughly, more data allows better learning. Again, quantitative

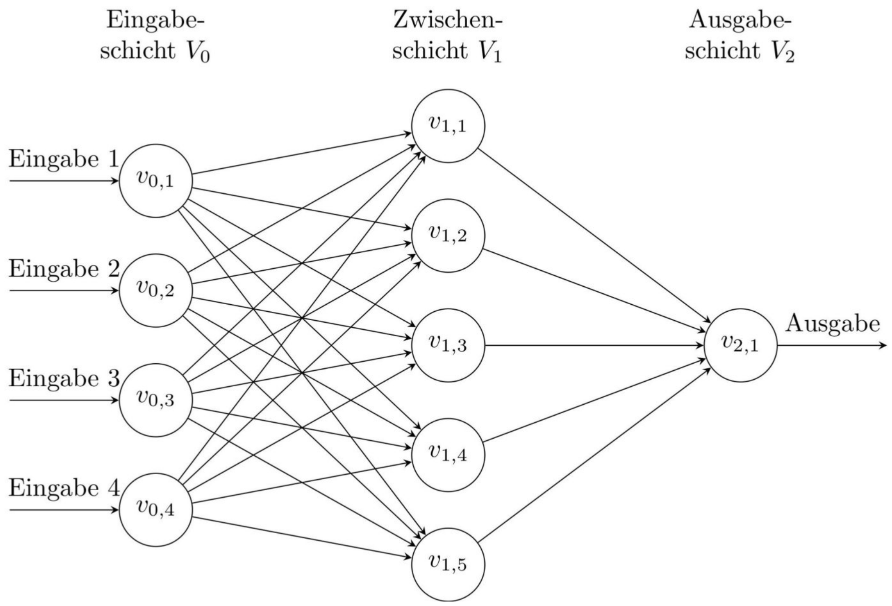


Figure 69: Sketch of a feedforward network with 3 layers. Deep neural networks in image processing are generally many orders of magnitude more complex.

answers are of interest. This question and finding ever more accurate answers is the subject of research in stochastics / statistics.

3. How are the arithmetic operations in neural networks implemented as efficiently as possible in computers? Engineers and computer scientists develop suitable hardware for this purpose, mathematicians and computer scientists develop efficient implementations (programs) of neural networks and also always new network architectures that make more accurate predictions and/or require fewer computing operations for this purpose.

48 Mathematical Architectures for Neural Networks***

You can't do AI without mathematics. Optimal network architectures for complex applications are best when they map the underlying mathematical theory. This is especially true for medical imaging applications.

AI, that's computer science, right? That's probably what most people think, although mathematics is also indispensable in this new, fascinating field of research. A particularly challenging field of AI applications are so-called inverse problems, including most medical imaging techniques as well as the parameterization of mathematical financial models or techniques for monitoring and controlling complex production processes. Prof. Dr. Peter Maass and his colleagues Sören Dittmer, Tobias Kluth, Johannes Leuschner and Maximilian Schmidt at the University of Bremen dealt with such problems.

But what exactly are inverse problems? These are problems where the physical or technical parameter is not measured directly, but is determined indirectly. A simple example is the measurement of the height of a mercury column (measurement) from which the current room temperature (sought parameter) is then 'calculated'. However, this example is too simple, in the narrower sense one speaks of inverse problems only if two difficulties are added. First, only processes are considered where arbitrarily small changes in the measurement can lead to arbitrarily large deviations of the sought parameter. For a long time such problems were called „badly posed", assuming, that the physical modeling was incomplete or inaccurate. It was not until the mid-1960s that methods of functional analysis were able to show, that this instability is unfortunately inherent in many problems and cannot be eliminated by better measurement techniques or more accurate modeling. This instability would not be a problem in itself if accurate measurement data were available. However, this is where the second difficulty comes into play: measurement data are never exact, but at best known within a given measurement accuracy.

In Figure 70 three reconstructions from CT data are shown, all of which reflect the given measurement data within the measurement accuracy of the tomograph. Now which is the correct reconstruction? One of the basic insights of inverse problem theory is that for such unstable problems, it depends on the particular application. In the present example, a physician making diagnoses based on tissue structure would prefer the middle one, while a physician interested purely in the shape of the organs would probably choose the right one.

Thus, there are no „correct" solutions, but at best „good" solutions, which are selected by a criterion individually formulated by experts in the application domain.

Such inverse problems are among the most complex challenges in applied mathematics, and their solution is also extremely computationally intensive. Initial attempts to solve this more efficiently using neural networks or other AI techniques, have failed or led to suboptimal results. A direct application of the AI methods that have been so

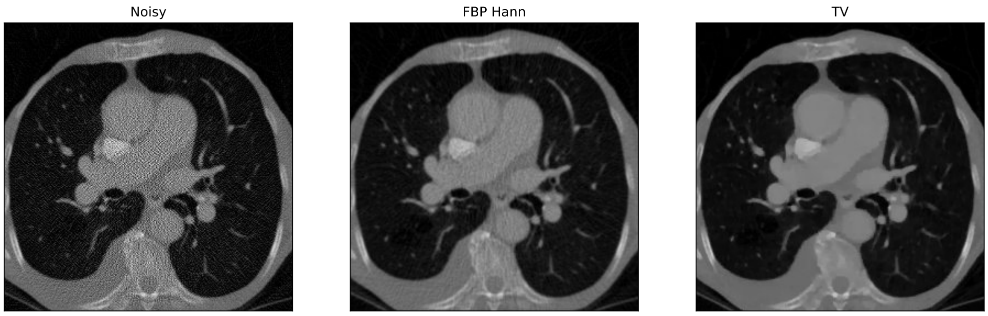


Figure 70: Three CT reconstructions generated from the same data set. The left reconstruction is obviously worthless (noisy), although it reproduces the measurement data just as well as the other two. Whether the middle or right reconstruction is preferred depends on what the medical expert wants to do with it. A diagnosis based on the tissue texture of the organs would prefer the middle one, a diagnosis a diagnosis based purely on the shape and morphology of the organs would prefer the right one.

successful in the field of computer vision and image processing, fails because of the inevitable instability of such inverse problems.

Before we talk about specific applications of neural networks to inverse problems, let's clarify, what neural networks are in the first place and how they are described mathematically. Neural networks are modeled after the human nervous system. In humans, stimuli are transmitted via nerve pathways and either amplified or reduced. In the so-called neurons, the incoming stimuli from different nerve tracts are then bundled together and transformed by a bio-chemical process into an outgoing stimulus for further transmission. Vividly, one can imagine the nervous system as a tangle of nerve tracts that transmit the incoming stimuli to a first layer of neurons. From there, the outgoing stimuli of this first layer of neurons are then analogously weighted and passed on to a second layer, and so on.

Mathematically speaking, the stimuli are simply numbers multiplied by weights (neural pathway) and then added together. The action of the neurons corresponds to the evaluation of a typically non-linear function, this is called the activation function. The result of this evaluation is then passed on to subsequent layers of neurons.

Ultimately, the result is a computational rule that receives a set of numbers as input and via the iterative application of weighting (linear transformation, corresponds to forwarding of stimuli along neural pathways) and activation function (corresponds to the actions that take place in a neuron) generates a fixed number of output values. A simple neural network is shown in Figure 71. Specifying the number of layers, the number of neurons per layer, and the choice of activation function determine the architecture of the network.

Just as humans learn from experience, a neural network can now be trained by presenting it with training data of known outcome. The weights of the network are then calculated in such a way that the network reproduces this result as best as possible.

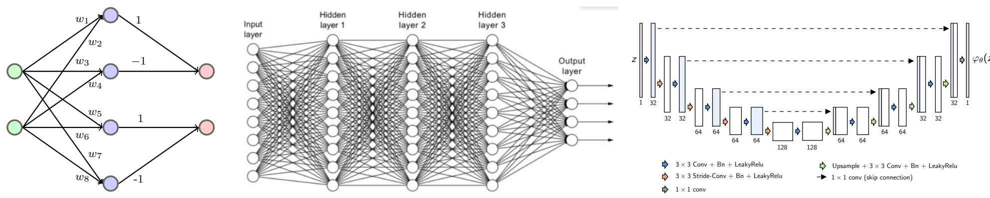


Figure 71: On the left is a simple neural network that receives two numbers as input and also outputs two values, the middle image is a „fully connected feedforward" network and the right image represents the commonly used „U-net" with about 500,000 free parameters.

To understand what this has to do with AI, one must first realize that neural networks, more precisely „fully connected feed forward" networks with L layers can be interpreted as the realization of L steps of an iteration procedure. Such a neural network has a vector x_0 as input variable. The values of the vector are transported from left to right, layer by layer. Between the layers, the individual values are multiplied with weights w, which corresponds to the forwarding incl. attenuation or amplification of stimuli along the nerve pathways. In the „neurons" of the individual layers, the individual incoming values are then combined and subjected to component-wise non-linear transformation, this corresponds to the biochemical reactions that in the individual neurons of the body e.g. decide whether the stimulus should be transmitted at all.

Mathematical Foundations

The way to a successful application of neural networks in inverse problems leads over network architectures, which are derived from deep research results on the analysis of inverse problems.

This requires insight into the mathematical foundations of both inverse problems and neural networks.

In an inverse problem – in a general formulation – an unknown parameter x, for which only measured values y^δ are known, which can be measured by a measurement process F with a measurement accuracy δ were measured data. In this formulation, all parameters x, which are

$$\|F(x) - y^\delta\| \leq \delta$$

satisfy potential solutions of the problem. Among these, the parameter that best satisfies an additional criterion R_{expert} is now selected. The exact specification of R can be quite general, here the application experts can encode properties, such as that edge information is particularly important in the reconstructions. Using Lagrange multipliers, this leads to a minimization problem

$$\min_x \|F(x) - y^\delta\|^2 + \alpha R_{\text{expert}}(x) \ .$$

Such minimization problems cannot be solved directly for a nonlinear F and nondifferentiable functionals R_{expert}. However, derivation by x and zero-setting allows the construction of iteration procedures which, starting from a suitably chosen x_0, approximate the solution step by step. Such gradient descent methods lead e.g. to

$$x_{k+1} = \text{Prox}_R\Big(x_k - \lambda F'(x_k)^*(F(x_k) - y^\delta)\Big).$$

To understand what this has to do with AI, one must first realize that neural networks, more precisely „fully connected feed forward" networks with L layers can be interpreted as realizations of L steps of an iteration procedure. Such a neural network has a vector x_0 as input variable. The values of the vector are transported from left to right, layer by layer. Between the layers, the individual values are multiplied by weights w, which corresponds to the forwarding incl. attenuation or amplification of stimuli along the neural pathways. In the „neurons" of the individual layers, the individual incoming values are then summarized and processed component by component, which corresponds to the biochemical reactions which in the individual neurons of the body e.g. decide whether the stimulus should be transmitted at all and if so how strongly.

Mathematically formulated, a fully connected feedforward network with L layers can be for solving the inverse problem can also be written as an iterative calculation. Given y^δ and with arbitrarily chosen x_0, $b = -\lambda F'(x_0)^* y^\delta$ is set. Then the application of the neural network corresponds to the iteration rule

$$x_{k+1} = \varphi\big(Wx_k + b\big).$$

For the special choice $W = I - \lambda F'(x_k)^* F(x_k)$, $b = -\lambda F'(x_k)^* y^\delta$ and $\varphi = \text{Prox}_R$ one sees the analogy with the analytic gradient descent procedure. Crucially, however, the activation function φ of the network is, according to the mathematical theory is chosen as the proximal mapping Prox_R of the expert information R. This would not have been thought of without the mathematical theory in mind. In any case, in this case, the output of the network, denoted $\Phi(y^\delta, W)$, corresponds exactly to the L-th iteration of the analytic gradient descent procedure.

The advantage of neural networks, however, is that the matrix W and the bias vector b are optimally fitted to the particular application using experimentally measured data: For this purpose, measured data y_i are collected for several known parameters x_i, $i = 1, \ldots, N$. Afterwards, the weights W are determined by minimizing the so-called loss function

$$\sum_{i=1}^{N} \|\Phi(y_i^\delta, W) - x_i(W)\|^2$$

is determined. Thus, the network can be optimally adapted to a given data structure and even finest modeling details of the measurement operator are included in the measurement data. This trims the network to produce outputs that are similar to those of the test data.

Application

As an application example, we will choose a particularly challenging and relatively new tomographic technique. In what is called „magnetic particle imaging“ (MPI), tiny nanoparticles are injected into the bloodstream. These nanoparticles have a magnetizable metallic core and a bioactive shell. These particles are now transported with the blood flow and allow – if their current position can be tracked – to provide detailed information about blockages in blood vessels and about cardiac function. In principle, the bioactive shell can also be constructed in such a way that the nanoparticles settle at certain metabolically characteristic regions and thus provide detailed functional insights. These nanoparticles are subsequently excreted via the natural circulation, making them potentially a harmless alternative to radioactive contrast agents.

However, the question arises, how can the position of the nanoparticles be tracked after they have been injected? To do this, a dynamic magnetic field is applied outside the body to magnetize targeted nanoparticles in specific areas. The dynamic change of this external magnetic field now causes nanoparticles to start oscillating and generate their own electromagnetic field. This in turn is measured by coils outside the body. In a simple linear model, the electromagnetic field, generated by a single nanoparticle in position x is denoted by $s(x,t)$. Now, if $c(x)$ is the number or concentration of nanoparticles at position x, the total signal is obtained by Summation (integration) of the contributions at all positions in the body Ω.

$$y(t) = \int_\Omega c(x)s(x,t)\,dx\,.$$

However, this is a very simplified model which, for example, ignores the complex measurement characteristics of the measuring coils and the particle-particle interactions as well as the influence of the size distribution of the nanoparticles. This is all very difficult to model. But, as mentioned, all these modeling subtleties are included in sufficiently good test data sets. This is what makes the use of data-driven methods in this area so attractive.

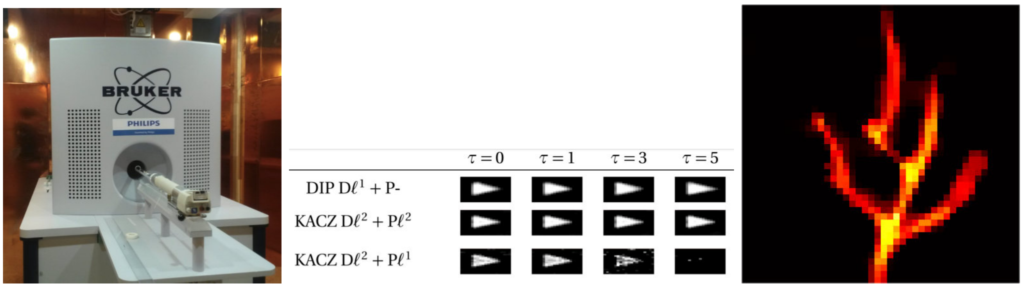


Figure 72: On the left is an MPI device from the UKE (University Medical Center Hamburg Eppendorf), the middle graphs are reconstructions of a real measured phantom, which was measured once conventionally (KACZ) and once reconstructed via neural networks (DIP), the right image shows an AI reconstruction of a vein tree.

49 How Social Networks Help Fix Software Bugs*

Mathematicians and computer scientists are using artificial intelligence methods to automatically categorize large numbers of bug reports in large software systems. Graph theory, as also used in social media, helps them perform this task particularly efficiently.

Bug Tracking in Open Source Software Projects

Software is used almost everywhere in our modern society, be it in our smartphones, cars, online services, or corporate production facilities. Their programming is usually so complex that it requires a team of software developers. Often, these are paid employees of a company that produces, tests and sells the software. Increasingly, however, software is also created on the basis of a so-called open source license by a community consisting of many volunteers who exchange information via online platforms and collaborate globally using Internet technologies. Popular examples of software created in this way include the Mozilla Firefox browser, the Android mobile operating system, and the software that forms the basis for the online encyclopedia Wikipedia.

Since there are often no paid test users in these so-called open source software (OSS) communities, user feedback is particularly important for the quality of the end product. For this purpose, almost all projects operate so-called „bugtracking„ platforms. This name is based on the fact that a bug in a software is also called „Bug„ (English for „insect„ or bug) in computer science - a term that comes from a time when computer processors were still so big that a bug crawling around inside could actually trigger a bug. If you as a user of open source software have a problem that you think is a bug, you can write a so-called „bug report„, i.e. a bug report. In it, you describe as precisely as possible which version of the software you are using, how the error is expressed, when and under what circumstances it occurs, and what result you would actually have expected. Now you hope that the developers will take a look at the problem and offer a solution as soon as possible.

Unfortunately, the sheer number of such bug reports poses a great challenge to the developers of particularly successful open source projects. For example, Mozilla Firefox developers had to process more than 112,000 bug reports from users over a ten-year period. If each of these reports took only 10 minutes to process, it would take a single person more than 2,300 working days. Interestingly, an analysis also shows that more than 80 % of the error reports are not due to an error in the software, but are caused, for example, by a misunderstanding on the part of the user or incorrect operation. So only a small part of this large number of bug reports actually helps to improve the software. If we can identify these helpful error reports as quickly as possible, this will help us to correct software errors as quickly and efficiently as possible. Can we use mathematical methods to solve this problem?

Social Networks in Open Source Projects

Nowadays, every child already knows social media like TikTok, Twitter, Instagram or Facebook. Even though these platforms differ in many details, they all have a common concept, which was developed in mathematics almost 300 years ago: a social network of users, which can be represented in the form of a so-called graph. You can think of such a graph as painting by numbers, i.e., it consists of points, called nodes, which can be connected in pairs by lines, called edges. Graphs of nodes and edges are used in many areas of science to represent complex relationships between different things, e.g. books in a library that are cross-referenced to each other, animal species in an ecosystem whose survival depends on each other, or people who communicate, know or are friends in a social network. The mathematical description and analysis of such complex graphs is the basis of many technologies known from everyday life. For example, search engines such as Google or Bing are based on an analysis of connections, so-called hyperlinks, between billions of web pages. Friend recommendations in social media such as Twitter or Instagram are calculated based on connections to our friends, as well as their connections to each other. And when we shop online, products suggested by shopping platforms are often based on a graph connecting those products that are frequently ordered together.

But how can mathematical analysis of such complex graphs now help manage the flood of bug reports in open source software projects? To do this, researchers have come up with a trick in which the members of an open source community are represented in the form of a social network. Whenever two people communicate with each other via the respective collaboration platform, they are connected by an edge. Over time, this creates a large complex graph that represents the community of volunteers of an open source project and allows us to analyze it using mathematical methods.

Whenever a user now sends in a bug report, not only is this bug report now available, but we also know the node in the social network that belongs to that user. How strongly this node is connected to other users, to which other nodes these connections exist, or at which position in the network they are located can now - in addition to the data of the bug report – be used by a machine learning procedure to automatically identify those bug reports that actually indicate an important bug in the software. The idea behind this is that a person's position within an open source community allows conclusions to be drawn about how well that person knows the details of the software and how much experience they have in describing a problem in such a way that developers can find the underlying bug in the code. So an artificial intelligence can now learn from which „regions"of the open source community's social network particularly helpful bug reports come.

A test of this method in a dataset with millions of bug reports on various open source projects shows that this method makes correct predictions in more than 90% of all cases, i.e. out of 100 bug reports identified by the method as potentially helpful, more than 90 actually result from a bug in the software. The method can therefore be used to quickly identify those bug reports that are most likely to actually result from a bug in the software. These can then be looked at first by developers in order to fix problems as quickly as possible. This procedure, which is based on mathematical concepts that are several centuries old, thus helps to find and fix software errors as quickly as possible in our digital society.

50 The Math Behind Netflix and Amazon Prime Recommendations*

An algorithm subtly suggests which book to read next or which movie to watch. There is linear algebra behind it.

Internet companies, such as Google, Amazon, Netflix, know best how to get to information they are looking for, such as a book or a movie. To do this, they use methods of linear algebra, that is, the mathematics of vectors and matrices, specifically matrix completion and *singular value decomposition*[48].

After watching a movie on a streaming service, you have the option to rate it with one to five stars. Based on these personal ratings and their selection of movies, it is determined which other movies you might like. Subsequently, one gets corresponding recommendations sent.

A simple way to find out if you would like, for example, „The Boat“ would be to average all the ratings the movie has received from other users. However, this is not a very good approach as it ignores whether you generally like submarine movies or not.

But how to decide which users are like you? This is where linear algebra comes in. Imagine a giant spreadsheet with columns named after each of the tens of thousands of movies available on Netflix. Each of the several million subscribers corresponds to a row in this spreadsheet. When a user rates a movie, an entry is filled in. For example, if I give „Avatar – Departure to Pandora“ three stars, the column corresponding to that movie is updated in the row with my name.

Few rate more than a hundred movies, so only a tiny fraction of this huge table is actually filled in. The task of the recommendation algorithm is to, fill in the entire table based on this very sparse information. How can this be done efficiently?

The Netflix algorithm

About three years ago, Netflix offered a million dollars to anyone who could improve the accuracy of its recommendations by 10 %[49]. Teams of software engineers and enthusiasts from around the world participated in the challenge. The most successful approaches are based on linear algebra: While the table of user preferences is huge, there are perhaps a few dozen stereotypical rating profiles. Each person's tastes are described as a mixture of these profiles. For example, you may like crime fiction and the occasional documentary. Linear algebra can identify these stereotypical profiles and provide the magic mix that describes your preferences.

[48] T. Hastie, R. Mazumder, J.D. Lee, R. Zadeh, *Matrix completion and low-rank SVD via fast alternating least squares*, The Journal of Machine Learning Research, 16(1) (2015), 3367-3402

[49] https://www.netflixprize.com/

Improvements of 8 to 9 % over the existing Netflix algorithm were achieved quickly after the price was announced. However, the required 10 % target proved elusive. After three years, a team of programmers from around the world finally cracked it in July 2009. They achieved this by taking into account that our preferences change significantly over time.

Yehuda Koren of Yahoo! Research Israel was one of the first participants to reach this 10 % mark by giving ratings a temporal component[50]. To do this, he had developed a model that could account for the temporal dynamics in the collaborative filters or recommender systems sensitively enough[51]. Although recent data reveals more about a user's current preferences, too much valuable information is lost by underweighting older ratings. Koren's trick was to retain everything that predicts the user's long-term behavior while filtering out the temporary noise, the "lunacy. To this end, Koren's model has one component for capturing the persistent signal and another for detecting signs of temporary noise, such as certain abrupt changes in ratings.

A Conversation with Machines

Powerful algorithms shape what you see in your web browser. They can also subtly suggest to you, what book to read next and what movie to watch. When you respond to these suggestions with your feedback, you influence not only your future recommendations, but also those of all other users of the same service. You are participating in a conversation between machines and humans.

[50] https://www.wired.com/2009/09/bellkors-pragmatic-chaos-wins-1-million-netflix-prize/

[51] Y. Koren, *The BellKor Solution to the Netflix Grand Prize*, (2009). https://www.asc.ohio-state.edu/statistics/dmsl/GrandPrize2009_BPC_BellKor.pdf

51 Taming Languages by Counting Words*

Thanks to a mathematical regularity in language, Google can answer a search query or automatically translate a text in a fraction of a second.

What is the most commonly used word in Dutch? That depends on what kind of language it is. When it comes to written Dutch in newspapers and magazines, „das“ is the most commonly used word. However, if you take spoken Dutch, the word „yes“ appears at the top. And on Twitter, „I“ is the top word.

But written Dutch, spoken Dutch, and Twitter Dutch all have one thing in common: the most common word within each such domain is twice as common as number two on the ranking, three times as frequent as number three, and so on. If we set the frequency of the most frequent word to 1, then the word frequencies form the series 1, 1/2, 1/3, 1/4, ...

This pattern expressed in a mathematical formula is called *Zipf's Law*, after the American linguist George Zipf, who discovered the law in 1935. „This law seems to apply to all languages and to all collections of texts within a language, whether you are looking up a Chinese law book, a Norwegian Bible, or English-language e-mails from a major corporation“, says Antal van den Bosch, a professor at Radboud University Nijmegen and a specialist in computational linguistics. „Zipf's law is an empirical law, but it's pretty accurate. Only at the beginning, for the top 10 words, and at the end, for the rare words, does practice deviate a little from the mathematical formula.“

Efficient Search

It is precisely because Zipf's law is universal that Google's search engine can respond so quickly. Van den Bosch: „The trick of Google is that they have created a word index of the web and are constantly updating it. The word index shows which word is in which document. Using Zipf's law, you can now show that this word index is compact. And that, in turn, means that you can store it compactly on disks and easily distribute it to data centers around the world."

Why exactly is this word index compact? Google has access to tens of billions of web pages, but the number of words per language is „only in the millions, of which there are usually only a few hundred thousand in an official dictionary. Zipf's law now teaches us that half of the words in a large collection of texts occur only once. Thanks to Zipf, we also know that the Top 300 contains almost all function words (articles, pronouns, prepositions ...) and the most frequently used content words (nouns, verbs, adverbs, adjectives). These two features make the word index compact.

Van den Bosch: „When we type in a search term, Google doesn't have to search through tens of billions of documents, but rather the much more manageable word index. And when someone enters four keywords, the search engine takes the overlap of four collections. Each set tells which web page the word appears on. It's a simple calculation, so it's quick."

Automatic Translation

Machine translation engines exploit a kind of derivative property of Zipf's law, namely the property of preventing word combinations. Google Translate uses a large database of existing translations, e.g. officially translated texts of the European Parliament or translated subtitles of movies. To translate a new text from Dutch to English, for example, the translation engine looks for the longest possible word combinations that have been translated in the same way as often as possible in existing translations.

For example, the translation engine sees that the Shakespeare quote „'Juliet is the sun" is always translated as „Juliet is the sun". In this case, it must be the correct translation. How often word combinations occur in a given order is also divided in Zipf-like fashion: only a limited number of combinations are very common. And just as the word index is compact, so is the index of word combinations. That is why a translation engine does its job so quickly.

This statistical translation method works well for texts that are very similar to existing texts. But the more unique and creative the text is, the more difficult it is for the translation engine. „Poetry is notoriously difficult", says van den Bosch. „The holy grail in my field is therefore: How can we ensure that machines really understand language? Because Google Translate still doesn't do that, no matter how useful it often is."

52 The Digital Twin – Rethinking Plant Operations**

Simulation is an essential tool in development today. Mathematical innovations allow simulation models to be reused during operation, and physical predictions to be made in real time.

Dr. Dirk Hartmann and Dr. Utz Wever from the Siemens AG, Technology, Munich summarize their experience with the Digital Twin:

What is a Digital Twin?

Digital twins are a new technology trend that already promises to have a major impact. They are the next wave of simulation technologies, merging simulation with big data and artificial intelligence technologies. Digital twins are so important to business today that they are one of the top 10 strategic technology trends for 2018. They are expected to become a business necessity that covers the entire lifecycle of a plant or process. For many products and services, Digital Twins will be an essential foundation. Companies that do not respond will be left behind. For example, companies that invest in digital twin technology are predicted to improve cycle times of critical processes by 30 %. The market potential for related offerings is predicted to be nearly 50 billion per year.

Digital twins collect all digital knowledge, models and data throughout the lifecycle of products and systems – from concept to end of life. They thus integrate model-based approaches, on which classical simulation and optimization paradigms are built, as well as data analytic approaches. Digital twins are powerful enablers of innovation and efficiency. They combine engineering knowledge, often in the form of executable simulation models, with available data and artificial intelligence. Through this integration, Digital Twins enable entirely new types of services, such as simulation-based monitoring and diagnostics or predictive maintenance, opening up new business opportunities.

To be a bit more concrete, we focus on one example, see Figure 73, which however occurs very often in practice.

What does Mathematics have to do with Digital Twins?

Mathematics is the language of Digital Twins. The blue boxes in Figure 73 symbolize the mathematical disciplines, which are needed for the construction of a Digital Twin. In the following, these will be briefly described.

The Model The basis of each Digital Twin is a computer model that describes the real system. This model can be represented, for example, by a neural network trained by real data. It is better if the real system can be described by physical equations (which is not always the case). For many applications, mature commercial computer programs

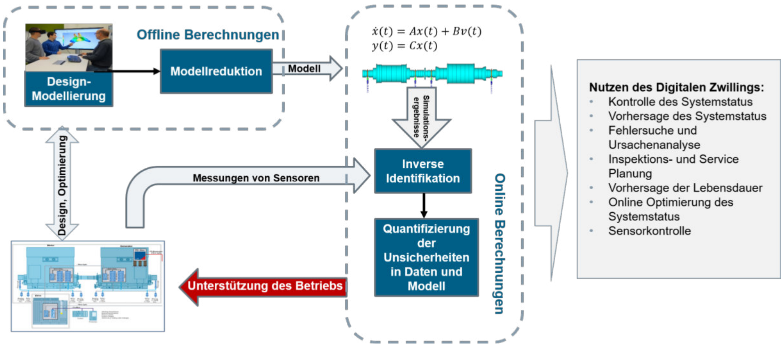


Figure 73: Digital twin to support the operation of plants, here using the example of the drive train of an electric motor.

already exist, which can be operated even by non-specialists. Major applications are structural mechanics, electromagnetism or fluid mechanics. Often, the modeling of increasingly complicated real systems requires a coupling of these computer programs. However, this leads to increasingly complex and lengthy calculations.

Model Reduction The core of a Digital Twin is a real-time capable model, which runs in parallel on a computer. The commercial computer programs described above are certainly too slow for this. Therefore, there are mathematical methods that generate a simple model from a complex model. The essential property of the simple model is to preserve certain physical properties of the complex model. This property can be well observed in the well-known *Stanford hare* (Figure 74). The right model is described with much less degrees of freedom, but it is still recognizable as a hare.

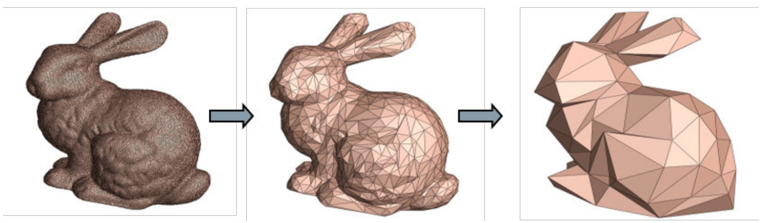


Figure 74: Model reduction using the example of the famous Stanford rabbit (from a lecture by Wil Schilders, Eindhoven University of Technology).

Inverse Identification The simulation results of the reduced model have to be compared with the sensor data of the real system, which mathematically leads to an inverse problem. Through this comparison, firstly, the co-occurring model is continuously synchronized, and secondly, many additional information about the real system can be obtained through deviations. A few examples are listed in Figure 73 in the right box.

Quantification of Uncertainties All data obtained by sensors and also the reduced model are subject to uncertainties. The resulting error cannot be reduced (only at high cost), but it can be quantified. Thus, a statement or result of the digital twin can only ever be made with a certain probability. If, for example, a prediction of the system status is desired, this can only be provided within a confidence interval. This phenomenon is well known from the weather map.

Example: A Digital Twin for an Electric Motor

The efficient thermal control of an asynchronous electric motor is considered as a use case.[52] Electric motors are subject to thermal material wear, which requires controlled temperature management. For example, large electric asynchronous drives are severely stressed by induction heating during start-up. Frequent starts without sufficient cooling pauses can lead to overheating of the motors. However, it is nearly impossible to measure the temperature of actual rotors spinning at high speed in the motor. Therefore, controls are often based on very conservative heuristics. In principle, appropriate temperatures can be calculated very accurately using 3D thermal simulations. In the case of the electric motors under consideration, linear convection-diffusion models are used. Corresponding 3D models are typically available from detailed design. However, they are computationally too demanding to be used in operation. However, appropriate models for background simulation can be realized by model reduction.

These are very efficient to evaluate and can not only be extended with uncertainty quantification methods, but also support continuous calibration with existing sensors on the stator side. Thus, with the help of continuously calibrated background simulation models, temperatures, which are not accessible for sensors, can be measured virtually. With the help of e.g. augmented reality devices, the user can virtually look inside the motor at any time and observe temperature distributions, cf. Figure 75 (a). This allows the cooling times of electric motors to be significantly reduced, which ultimately increases the plant availability. Enriched with methods for uncertainty quantification, confidence intervals for the rotor temperature can be provided (cf. Figure 75 (b)), which allow to go close to operating limits.

[52] D. Hartmann, M. Herz, M., U. Wever, *Model Order Reduction – A Key Technology for Digital Twins*, KoMSO Challenge Workshop, Springer (2017).

(a) (b)

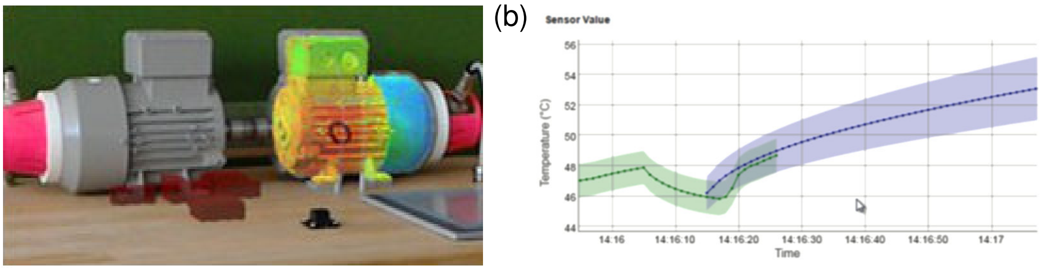


Figure 75: (a) Temperature visualization inside an electric motor using augmented Reality using the Microsoft Hololens.
(b) Screenshot of the virtual temperature sensor solution for estimating the current rotor temperature (green) and predicting the rotor temperature after startup (blue) at a specific point. The plotted bands show the calculated uncertainties, cf. https://www.youtube.com/watch?v=86vkjykbHRM

53 Taming Financial Markets with Math**

A financial market is like a living organism, which is continuously changing, unpredictable, and yet can be tamed thanks to stochastics.

Dr. Martin Friesen of the School of Mathematical Sciences, Dublin City University, Ireland, describes his experiences on financial mathematics:

While share prices in the world's major financial centers were still at record levels at the beginning of 2020, the mood among investors turned abruptly when the Corona pandemic took a stranglehold on the world and turned our daily lives upside down. Corona-related lockdowns temporarily paralyzed our economy and, as a consequence, the financial markets shorted out worldwide in February/March 2020. Although it was possible to stabilize the financial markets again through, among other things, rapid and spirited intervention by the central banks,[53], the *Corona Crash 2020* impressively demonstrates that unforeseeable events can occur at any time and are associated with non-negligible risks. It is therefore imperative that all major and, in some cases, systemically important players in the financial markets (such as banks, insurers, but also sovereigns) hedge against a wide range of possible risks.

Stochastics, a branch of mathematics, is needed to counter such uncertainties. This makes it possible, for example, by means of the probability calculation and statistics, to to continuously analyze possible future scenarios on the basis of past events and to create forecasts for the future by means of mathematical models. The currently largest, though not only, system for fully automated risk analysis is called Aladdin[54] and is operated by the New York financial services provider Black Rock[55]. In addition to large global corporations, these service providers also have the European community of states as customers.

Understanding Risks

But why should actors act on the financial markets at all if they are associated with risks that must first be tamed with complex mathematical models? This question can be answered most simply with the saying *„The money must work"*. So it is the possibility to participate in the success of companies through investments (e.g. through dividend payments or rising share prices). The exact connection of profit and risk is thereby described by the *main theorem of financial mathematics* and can be abbreviated with the

[53]https://www.bundesbank.de/de/aufgaben/geldpolitik/geldpolitische-wertpapierankaeufe/pandemic-emergency-purchase-programme-pepp–830356
https://www.faz.net/aktuell/finanzen/finanzmarkt/ezb-stockt-krisenprogramm-pepp-um-500-milliarden-euro-auf-17095348.html

[54]https://www.it-times.de/news/aladdin-der-super-algorithmus-von-blackrock-135395/

[55]https://www.faz.net/aktuell/finanzen/fonds-mehr/vermoegensverwalter-blackrock-der-schwarze-riese-12057048.html

saying *„There is no free lunch"*. Excluding interest, this relationship can be expressed in a single mathematical formula as follows:

$$\mathbb{E}_u[R_{t,u}] = 0, \qquad 0 < u < t. \tag{1}$$

In concrete terms, it means that the average profit or loss achieved at time t, provided one has bought a security at time u, is always zero. The easiest way to understand this formula is to use a simple coin toss as a random experiment. If heads is tossed, the stock price rises, but if tails is tossed, the stock price falls. Each individual coin toss is independent of the others (one says *the coin has no memory*). Since we expect heads and tails to occur equally often on a fair coin, the average outcome is zero. However, this says nothing about the outcome of a single investment (just as one cannot predict a single coin toss). So there is definitely the possibility of making a higher profit, but in return you have to be willing to accept possible losses. So in short: *There is no such thing as a surefire bet*.

Computer Trading: Many Opportunities and New Risks

If about 50 years ago a financial market was still a purely physical marketplace where purchases and sales were made by real people, this has changed fundamentally since the 2000s at the latest. Today, financial markets are thoroughly digitized, automated and globalized. In fact, the overwhelming majority of all buying decisions are already made by sophisticated algorithms, which act and react in fractions of a second[56]. While such computer-driven systems are based on mathematical models on the one hand, they are also increasingly equipped with artificial intelligence, which, it is expected, automatically adapts to new situations and always makes the best rational decision.

One consequence of modern „computer trading" is, that price developments show very high swings in short periods of time. In financial mathematics, this is referred to as volatility.

Current research work in mathematics is therefore also concerned with the question of how such effects can be captured as precisely as possible. The approach with so-called *„Rough Volatility models"* has proven to be extremely effective. An important feature of these models is the An important feature of these models is the addition of earlier price fluctuations according to the motto „strong fluctuations lead to strong fluctuations again in the short term„. With respect to the coin flip described earlier, here the coin would additionally get a kind of „memory„ assigned, so that for possible predictions one must also always consider the past price swings. Specifically, results already available today can be used to calculate *fair prices* for hedges against price losses. In this way, financial market players cannot eliminate their risks entirely, but they can at least control them and thus *tame* the unpredictable financial markets.

[56] https://www.heise.de/newsticker/meldung/Hibernia-Express-Erstes-neues-Transatlantik-Kabel-in-12-Jahren-2778847.html

Figure 76: DAX March 17, 2021

54 Modeling Negative Interest – Only Cash is True?**

Mathematical models help to model current developments in the financial markets and can help to gain insights by means of economic models.

Jörg Kienitz of Acadiasoft Inc, is a private lecturer at the University of Wuppertal, Adjunct Associate Professor at the University of Cape Town and operator of the site finciraptor.de. He uses the example of negative interest rates to describe how mathematics helps to model current financial market phenomena.

Negative interest rates are now well known to all of us. Ultimately, we learn this when we compare interest rates on overnight money or read about „custody fees“ for deposits. How did this situation come about? In the wake of the 2007 financial crisis, the European Central Bank needed to maintain stability in the Eurozone. To do so, it relied on increasing the monetary base by buying government bonds („quantitative easing“) and lowering the key interest rate. The latter determines the interest rate at which commercial banks can borrow money. However, the gradual reduction that was initiated did not have the hoped-for effect, which led to further reductions and eventually to this interest rate becoming negative. Currently[57] it stands at -0.5 %.

Countries such as Switzerland, Sweden, Denmark and Japan have already applied negative interest rates. In the case of Switzerland, this move resulted from strong money inflows, which were invested in Swiss interest-bearing bonds for lack of alternatives. The impact on exports or even the increase in the cost of consumption was considerable. The Swiss National Bank countered this by lowering the key interest rate below 0%.

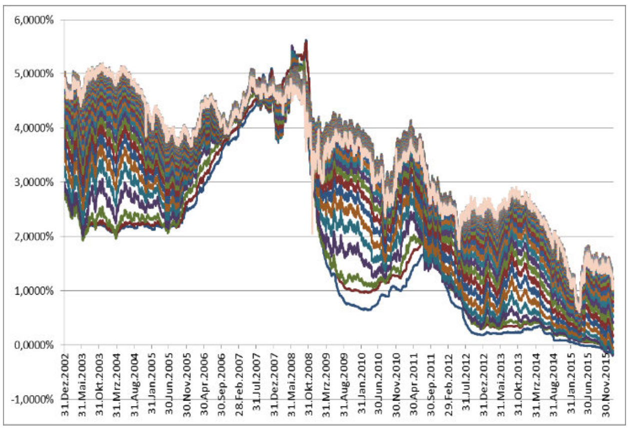
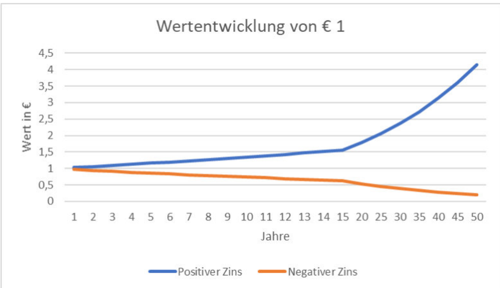


Figure 77: Development of different interest rates (swap rates), EUR in the period from December 31, 2002 to November 10, 2015.

[57] as of April 14, 2021

In mathematical modeling, we consider interest rates, which are linked to a maturity. Interest rates are quoted on the market at different maturities, see Figure 77. An interest rate describes the development of a money supply, when it is invested for the corresponding term. From such interest rates, yield curves are constructed. Thus, in Figure 78, yield curves for constant interest rates of 3 % and -3 % are shown. In addition, the so-called discount curve is shown. While one can read from the yield curve how 1 € develops for different maturities, the discount curve shows the amount of money needed to obtain 1 € at a given maturity, see Figure 78. For positive interest rates, the value is smaller than 1, but for negative interest rates, it is larger.

For the determination of the value of an interest rate transaction, a distinction is made between products which are determined solely by the current interest rate curve. Thus, the value of an instrument which pays an interest rate of 3 % every year, e.g. on June 11, and after 10 years also pays the nominal value, e.g. 100 €, can be calculated. For this purpose, one reads the corresponding value in the discount curve, multiplies it by the interest rate (nominal value · 3 % = 3 €) and adds up these values. Finally, the discounted repayment value is added. For the discount curves from Figure 78 exactly 100 € results, while in the case with negative interest rates 171,37 € is calculated. Such instruments are also called linear. In addition, there are also non-linear instruments which include, among others, interest rate limiters. The payoffs of such instruments are non-linear functions of an interest rate, e.g. the maximum or minimum of a limiting value.



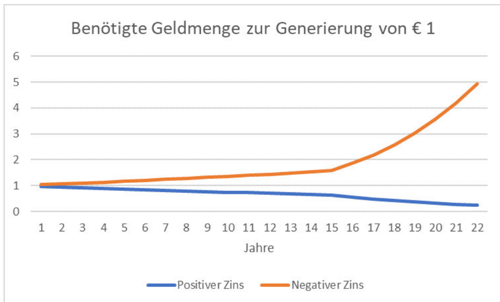


Figure 78: left: Yield curve for a constant interest rate of 3 % (blue) and -3 % (orange); right: Discount factor for a constant interest rate of 3 % (blue) and -3 % (orange).

It is not possible - not even in the linear case - to calculate this value for all future points in time. The yield curve is a dynamic object driven by trading in money and capital markets as well as decisions of commercial and central banks or investments of pension funds and funds. This is shown in Figure 77, where interest rates are shown for different maturities in the period from 12/31/2002 to 11/10/2015. Both the amount and the ratio of interest rates to each other fluctuate.

Therefore, dynamic, stochastic models are used in financial mathematics to describe developments in interest rates. Prices of interest rate products are then calculated as expected values with respect to probability distributions. For interest rate products it is guaranteed that the modeled interest rates are hit exactly. Now we see why products that require only the current interest rates are called linear. This follows from the linearity of the expected value. For non-linear functions, the calculation is not so simple. Here the knowledge of the distribution is needed to be able to determine the value. In this case, other parameters such as volatility, skewness or kurtosis also play a role in determining the distribution.

This stochastic development of interest rates is very complicated. It is possible to model one interest rate, several interest rates, or the entire yield curve, i.e., infinitely many interest rates simultaneously. In the simplest case, we model the stochastic evolution of a stochastic factor and consider the curve as a function of this factor. The evolution is described by a deterministic part, the so-called drift, and a stochastic part. The latter is often assumed to be Brownian motion. In this case, the stochastic part is called diffusion.

The dynamics can be controlled in various ways by a clever choice of the diffusion component[58]. For example, this component can be modeled as a multiplication by a constant and the random realization of Brownian motion but a function of current state and time. If one wants to use a normal distribution, one chooses the first approach. One also says the stochastic motion reacts absolutely to the random motion of the driving Brownian motion. If, on the other hand, one chooses the second approach, one obtains a log-normal distribution, see also Figure 79 (left panel). It is interesting to note that this choice also restricts the domain of definition of the stochastic dynamics. In the first case, any real number can be realized. If one chooses the second option, this is no longer possible and only positive real numbers can be realized.

How can we now turn to the modeling of interest rates? At the beginning of their modeling, simple models based on normal distributions were used. At that time, however, negative interest rates were an artifact and resulted from the simplicity of the models. However, care was taken to use, if possible, model parameters that allowed only a very low probability of negative interest rates being realized. Later, these models were further developed to avoid negative interest rates. At the same time, more or infinite dimensional models were considered to represent ratios of interest rates to each other. The focus was rather on other market conditions, e.g. the so-called skew. This describes the fact that for the calculation of non-linear products the properties like skewness or kurtosis have to be considered, see Figure 79 (right side). For this purpose, it was convenient to use mixtures of normal distribution and log-normal distribution. This leads, for example, to so-called shifted diffusions. These tools allowed a better modeling of the skew, but again resulted in the possibility of realizing negative interest rates.

In the current market situation, one again resorts to such models which explicitly allow negative interest rates. Certainly, the interest rate cannot assume arbitrarily small values, which makes the use of the normal distribution seem questionable. Ultimately,

[58] The drift is determined by the diffusion component and the current yield curve for reasons of financial mathematics.

however, one can again get out of the loop with the choice of parameters and try to let probabilities below a threshold become very small. Here, the shifted diffusions allow an explicit, direct modeling of the threshold. In these cases, negative values can be modeled up to the threshold, see Figure 79 (right panel).

It should be noted that the stochastic models and numerical methods used are unproblematic with respect to negative interest rates. However, significant problems arise in economic model building. For example, consider the threshold. A theoretical lower bound can be determined by the existence of cash, which then depends on withdrawal fees and storage costs. In addition to this physical lower bound, attempts are also made to establish more economically plausible values by means of more complex models. For the cash case, assuming withdrawal fees of 2 % and storage costs of 1 % per year, the lower bound is -2.98 % per year, - (100 % - 2 %) · (100 % - 1 %) = - 0.98 · 0.99 = -2.98 % per year. Economic difficulties and implications of negative interest rates can be considered mathematically in terms of models, and (complex) numerical and stochastic methods are available to implement these models. However, these models are based, among other things, on economic and political assumptions, which, of course, cannot be verified mathematically. Here, a close interaction of the different disciplines is necessary. Since the subject matter has implications for societies and ultimately people should be handled with great care and open communication and presentation of assumptions. Ultimately, everyone must answer the question posed in the title for themselves. However, cash can be used to establish a lower bound for negative interest rates, which would not hold without „cash“.

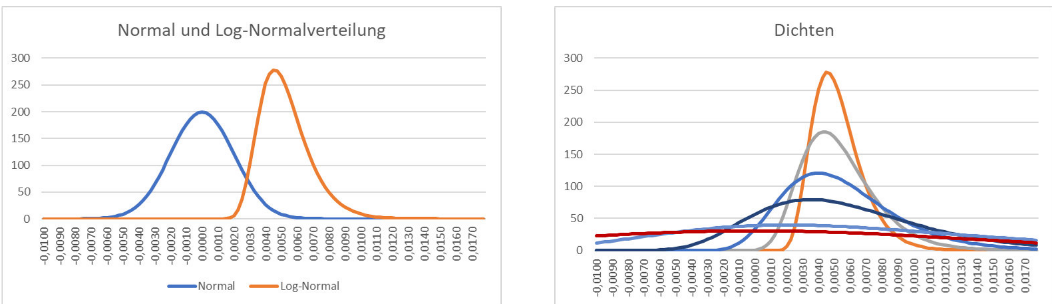


Figure 79: Left: Normal and log-normal distribution; right: distribution to log-normal (orange), shifted diffusions with different shift parameters and normal distribution (red).

55 Insert for Financial Instruments: PRIIPs!**

How to compare the return prospects of different financial instruments? In an ideal world for women investors, wealth increases steadily, without much excitement. In reality, however, different investment instruments have vastly different maturities, risk and return profiles and are therefore ex ante, i.e., before an investment decision is made, very difficult to compare.

For the so-called *Packaged Retail and Insurance-based Investment Products (PRIIPs)* the European Regulation (EU) No. 1286/2014 stipulates that each such PRIIP must be provided with a Key Information Document (KID) in which the manufacturer (usually a bank, insurance company or investment company) quantifies the risk, potential return and costs associated with the product in a standardized manner on no more than three A4 pages. This is intended to make different PRIIPs comparable.

What are examples of PRIIPs? What are examples that are not PRIIPs? The key words are „Packaged" and „Retail": When a (for example: floating to fixed) Swap transaction[59] with an institutional investor, then it is not a PRIIP, if a swap with exactly the same terms is sold to an end customer (retail), then it is.

Similarly, if an instrument is not packaged, it is not a PRIIP. An individual share of Apple is not a PRIIP, a German federal bond is not a PRIIP, a Euribor-rate issue by Commerzbank is not a PRIIP.

In contrast (always assuming the retail client): an exchange traded fund on the EuroStoxx50 is a PRIIP (packaged), a callable fixed rate bond of Credit Agricole is a PRIIP (bond packaged with call right), a floater with cap and floor of Hessische Landesbank is a PRIIP (floater packaged with caplets and floorlets).

Likewise, PRIIPs are (exemplary, in no way exhaustive list): swaps, options, futures, securities funds, convertible bonds, structured equity bonds, structured interest rate instruments, unit-linked life insurance.

Where does the Math Hide?

In addition to the PRIIPS Regulation, the Delegated Regulation (EU) 2017/653 prescribes how exactly the basic information sheets are to be designed and which calculations are to be made for this purpose. For the majority of PRIIPs, the calculations to be made for risk quantification and the performance scenarios (earnings prospects) are the most complex. Such valuations calculate the fair value of a financial instrument based on currently available market data, but for PRIIPs that have not yet expired, that data is in the future. In undergraduate mathematics courses, such valuation techniques are covered in „Financial Mathematics".

[59] such a swap exchanges fixed and floating interest payments for a fixed notional amount. For example, for a home financing, the interest rate of a loan can be converted from fixed to floating (or vice versa) without having to terminate the base loan at potentially high cost

Andreas Binder of UnRisk has been working with his team of mathematicians, physicists, and computer scientists for more than 20 years on mathematical software for the valuation and risk analysis of financial instruments. On PRIIPs, he notes: „The idea that different instruments are made comparable is, in my view, a right and important step towards investor protection and cost transparency. In detail, however, the preparation of a basic information sheet can be enormously time-consuming. According to the Delegated Regulation, 70,000 valuations must be carried out for long-term (> 5 years) structured interest rate instruments. In this context, valuation means solving a partial or a stochastic differential equation with possibly tens of thousands of unknowns on the time-interest rate grid. If a single valuation takes only one second, we are talking about computation times of one day on the single CPU."

Together with the Technical University of Berlin, the Austrian company MathConsult is working on model reduction techniques for risk analysis in a subproject of ROMSOC (www.romsoc.eu).

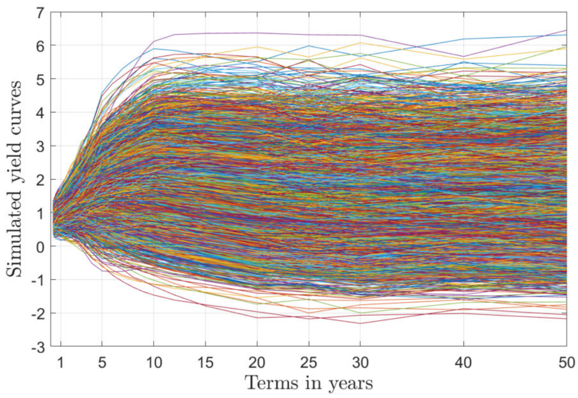


Figure 80: For the year 2031 (ten years from now), the 10,000 simulated yield curves in the figure. Each individual curve shows on the y-axis the interest rate for the term (in years) on the x-axis.

56 Calculation of Camera Images**

Unmanned aerial vehicles use camera images for navigation. But there are all kinds of distortions in these images: Angles change, objects deform. It takes a lot of trickery to calculate what reality looks like.

Camera images are used in a wide variety of places. Jaap van de Loosdrecht and Klaas Dijkstra of the Computer Vision Knowledge Centre (NHL University of Applied Sciences) are working on unmanned aerial vehicles, for example, on a project to monitor the movements of cows in a barn. There are several cameras in this barn, and the intention is to indicate on a map where each cow was at what time. This is more difficult than it seems. The cameras are all hung at a different angle, and a cow may look much larger in one picture than in another. The wide-angle lenses used also result in barrel distortion: everything in the picture seems to be pulled to the edge. All in all, the images are distorted in all sorts of ways.

To map this distortion, the researchers create a pattern in front of the camera from which all distances are known. They use a pattern with ten rows of fourteen dots each. All the dots are black except for the first two in the top row, so you can always see what the top left corner of the image should look like. From the distortion of these neatly arranged dots, you can deduce the exact image distortion. This is called camera calibration. The purpose of calibration is to eventually calculate from a camera image back to the original three-dimensional coordinates in reality.

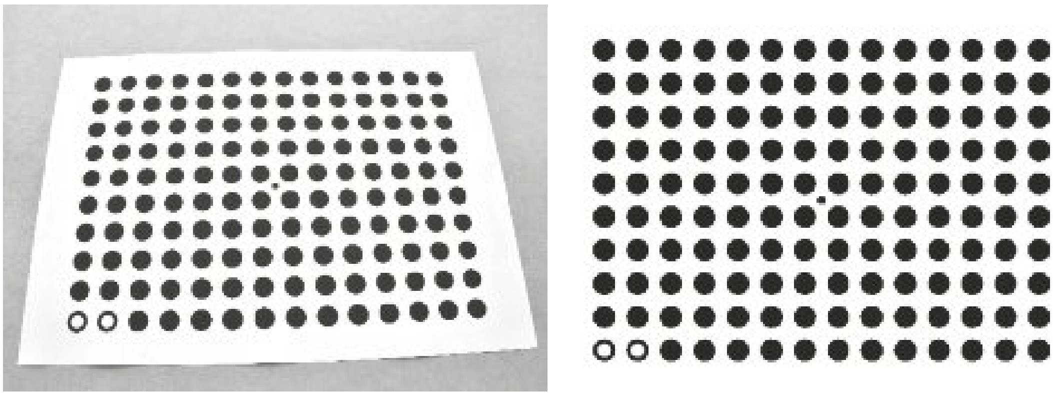


Figure 81: On the left, the dot pattern; on the right, the photo it creates. The lines are no longer straight, and some dots appear larger than others.

Point Cloud

Fifteen unknowns must be estimated during calibration. These are e.g. the distortion of the lens in three different directions, but also the angle at which the camera is hanging. We use this pattern of 14×10 points. So you get a cloud of 140 points in a 15-dimensional space and look for a nonlinear relationship between these points.

The researchers describe their view of mathematics as practical: they are particularly interested in solutions that work well. Therefore, they are not looking for the exact solution, but for an approach that is good enough for their applications. That they do with so-called *genetic algorithms*: a computer program that mimics evolution. They start with a number of random estimates for the 15 unknowns. Then, for each of those estimates, they calculate the corresponding error: The better the nonlinear relationship found describes the point cloud in 15-dimensional space, the smaller the error.

Next, the best estimates are allowed to reproduce themselves (it is simply survival of the fittest, „survival of the fittest"). Two estimates are combined to form a new estimate. As in nature, its „offspring" is not only a mixture of its two parents, but a few random changes also take place. This creates a new generation. This process is repeated over and over again. Each time the best estimates are allowed to reproduce, until a solution is found where the error is small enough. How small the error can be depends on the application.

Aircraft

Unmanned aerial vehicles must be able to estimate pixel-precise distances and angles from their camera images. In the future, these small aircraft will have to perform inspections of wind farms, instead of the humans who now do so at the risk of their own lives. Aircraft used now often have extra cameras and need extra computers to do on-site calculations. Efforts are being made to install cameras and computing technology in the aircraft itself, because only then can this technology really be deployed everywhere. But to do that, we need to be able to analyze these camera images very well.

57 Mathematics Proves Software has no Bugs*

Software bugs can cause hundreds of millions of dollars in damage and even cost lives. Mathematics helps to prove that the software is absolutely error-free.

As computer software becomes more and more complex, it becomes increasingly important to reduce the probability of errors as much as possible. This is certainly true for life-critical software applications in cars, airplanes and hospitals, but increasingly also for enterprise software. The Dutch company ASML is the world's largest manufacturer of machines that print computer chips on silicon wafers. Major chipmakers such as Intel, Samsung and TSMC use ASML machines to print their own computer chips on silicon wafers. These chips are found in the latest iPhones and iPads, for example.

Each ASML machine is controlled by a colossal software program. The foundation for this computer program was laid 25 years ago and has been continuously expanded and improved ever since. The program now has more than thirty million lines of code, and no one can miss all the details. ASML employs 900 people to maintain, improve and expand the software. Software errors can cause hundreds of millions of euros worth of damage and even cost human lives. Mathematics helps to prove that a piece of software is absolutely error-free.

Costly Bugs

Programming is a meticulous job and a golden rule in the software world states, that on average there are 10 errors in 1,000 lines of computer code. For the ASML machine, this means that there can be as many as 300,000 errors in the software. In practice, a customer will not notice many of these errors, but some errors can shut down the machine for hours. The machine costs 40 million euros, and for every hour the machine is stopped, ASML customers lose who use the machine quickly lose hundreds of thousands of euros in revenue.

Traditionally, bugs are discovered by testing software. The problem with testing is that you can prove the presence of bugs, but not that it does not contain bugs. Because software bugs are so expensive, ASML has been using mathematical proof techniques for several years, that can prove that parts of the software do not contain bugs.

Every piece of software is essentially a sequence of decisions: if A is true, execute B; if A is not true, execute C. Assuming a program contains one of these types of decisions, it can be in 2 possible states. With 10 decisions, this is already $2^{10} = 1024$ states, and with 1000, it is even up to 2^{1000} states. If one wants to guarantee that there are no errors in the software, then one must try all possible combinations of decisions. Even with a hundred combinations per minute, it is not feasible for a large system like ASML's, to achieve this in a reasonable time frame.

Mathematical Tricks

The trick is to reduce the number of possible states that software can be in. Suppose the program has to do tasks A, B and C and the order does not matter. Thus, all six combinations ABC, ACB, BAC, BCA, CAB and CBA all end up in the same state Q.

Figure 82: Example of an ASML lithography machine. Some major challenges for ASML lithography machines: very accurate, ultra-fast and ultra-small line writing on silicon wafers that are to be turned into computer chips. Source: ASML

In the classical testing of the software you have to try all six combinations. The mathematical method of proof sees, that all combinations lead to Q. In this method of proof, you only have to think about a much smaller number than the total number of possible states. Then it is possible to prove the absence of errors in parts of the whole software.

Another trick is the *symmetry reduction*. Suppose the ASML machine can process 3 products at the same time, while 6 products can be in the machine at the same time. The mathematical proof method then sees, so to speak, that the processing of product

1, 2, and 3 is equivalent to is equivalent to processing product 2, 3, and 4, and so on. However, the engineer must convey this knowledge by telling the mathematical proof method, that the products are processed neatly in order. Ultimately, such mathematical tricks can be used to prove that the robot arms moving the silicon wafers, never collide, or that the order in which measurements are taken is always correct.

58 Understanding Crime through Math***

Why do people commit crimes? Why are some unfazed by the consequences, while others would never break the law even when no one is looking? Are there environmental, socioeconomic, or even biological factors that lead to criminal behavior? Are there possible deterrents? What is the appropriate punishment? How do we recognize wrongdoing in the first place? Can all these questions be put into mathematical terms?

Dr. Yao-Li Chuang and Prof. Dr. Maria R. D'Orsogna, Department of Mathematics, California State University, Los Angeles, summarize their work on modeling crime:

Some crimes are known to happen in the heat of the moment. They are driven by sudden and strong impulses, triggered by jealousy, anger, or in response to a provocation. These „crimes of passion" are often explained as cases of temporary insanity. Premediated crimes, on the other hand, involve planning and careful research beforehand. Emotions such as revenge, greed, or envy may still motivate the crime, but planning is done in advance for how the dynamics will unfold, often with the intent of evading prosecution. Finally, crimes of opportunity occur when an individual recognizes that circumstances are optimal to commit an illegal act and decides to seize the opportunity without being provoked or without intent.

Residential burglaries are representative examples of crimes of opportunity: unless there is a specific intent to rob a place where valuable items such as money, art, or jewelry are kept, residential burglaries are the result of three converging factors: an easy/valuable target, the absence of a guard, and a motivated offender. Simply put, a burglarized home is in the wrong place at the wrong time when the burglar strikes. Had the owner been present or had the security system appeared more secure, the burglar would have chosen a different location or perhaps none at all. First introduced in 1979 by Lawrence Cohen and Marcus Felson, routine activity theory formalizes the emergence of opportunistic crime in sociological terms and has been applied to burglaries, sexual assaults, pickpocketing, vehicle thefts, and robberies.

Evidence suggests that repeated or near-repeated offending is typical in crimes of opportunity. In the case of burglaries, this means that the same house (or neighboring houses) is repeatedly burglarized. Although it may sound counterintuitive, repeated and near-repeated crimes occur because the offender is familiar with the location, routines, and security after the first offense, and returning to the same house (or nearby houses) allows for less uncertainty and more efficiency. The original offender may also have passed on information to others, so imitation by another offender may also play a role. Repeated and near-repeated burglaries have been reported in many urban centers worldwide, with the risk of burglary increasing over several weeks and several city blocks after an initial burglary. Some statistical measures show that prior crime is the best predictor of future crime.

'Broken windows' are a related and well-known theory in criminology that postulates that visible signs of even minor crimes, if left unattended, encourage further and more serious crimes. The theory was developed in 1982 by James Wilson and George Kelling[60], and the title refers to the sight of broken windows that were so ubiquitous in New York City in the late 1970s, a place and time marked by high crime rates. Signs of disorder in the neighborhood (broken windows, as well as graffiti, trash, parked cars without license plates) send the message that the area is not well-maintained or well-guarded, and that illegal activity is tolerated. This message attracts more and worse crime. Soon, one broken window or graffiti can become many, signaling urban decay. Since the late 1980s, this theory has informed policing in New York City, which developed a zero-tolerance strategy for even minor crimes as a deterrent to major crimes. The experience of other communities in Massachusetts and the Netherlands shows that increasing the sense of order in neighborhoods helps curb crime. However, the broken-windows theory has become more controversial in recent years because it can lead to excessive preventive interventions and excessive policing, as in the use of „stop and frisk" in New York City, in which people suspected of being criminals because of their appearance or behavior are briefly stopped by police without any real evidence of involvement in a crime.

Opportunity crimes that incite further crime are best treated mathematically because they do not involve individual, highly personal motivators as in revenge killings or love triangles. Residential burglaries, in which offenders repeatedly visit the same locations and create „hotspots" of crime, are a natural starting point because the target, a house, is fixed and only the movement of the offender needs to be described. The goal is to create a mathematical framework that incorporates elements from the sociological and criminological theories described above so that one can study burglaries (and other opportunistic crimes) with quantitative tools.

We begin with an agent-based model in which a burglar wanders through a virtual city, e.g. on a grid, and takes a random walk (English, 'random walk') that is influenced by a dynamically evolving „attractiveness field". Each time a crime is committed, the burglar is removed and later reappears in a different location, ready to strike again. Once a crime is committed, the attractiveness of the location increases. The idea is that after a crime is committed, the offender returns to his or her home location and that the increased attractiveness directs the mobility of future offenders toward previously burglarized locations through a positive feedback loop, according to the principles of repeated and near-repeated offending. Attractiveness will also naturally decrease to an intrinsic baseline value when no crimes are committed.

Under certain conditions, these rules can be cast into a compact system of so-called diffusion-reaction differential equations that couple and account for A, the density of

[60] J.Q. Wilson, G.E. Kelling, *Broken Windows. The Police and Neighborhood Safety*, in The Atlantic Monthly. March 1982

criminals and the attraction field, as follows

$$\frac{\partial A}{\partial t} = \eta \nabla^2 A - (A - A_0) + \rho\, A,$$
$$\frac{\partial \rho}{\partial t} = \vec{\nabla} \left[\vec{\nabla} \rho - \frac{2\rho}{A} \vec{\nabla} A \right] - \rho\, A + \Gamma.$$

Here, the dispersion parameter η represents how much the attraction to near A_0 is the base value of the virtual city's attractiveness to which A naturally gravitates in the absence of crime, A_0 is the base value of the virtual city's attractiveness to which A naturally gravitates in the absence of crime, and gamma is the rate of arrival of criminals to the city. Hotspots occur when crimes cluster in certain areas due to a sharp increase in the attractiveness field that attracts criminals after a burglary and are sustained by large numbers of criminals visiting those areas. However, when criminal activity is too high, e.g., when there are too many criminals and the attractiveness field becomes uniformly high everywhere, the hotspots merge and eventually the entire city becomes a single crime desert. The results for the attractiveness field A for different parameters are shown in Figure 83 and show where criminals tend to cluster and where burglaries are more likely. This model has been extended to include spatial disorder in the city (by accounting for non-uniform values of A_0) and, most importantly, to explore methods for weeding out hotspots and determining the strategic location of law enforcement agencies.

The above phenomena of clustering of crimes and near-recurrence of events are also observed in gang crime, terrorist attacks, and protests, due to other forms of contagion in which an initial random event may trigger other nearby events, either as „imitators“ or retaliation. The temporal patterns are reminiscent of earthquake activity, where an initial event triggers many aftershocks nearby. As a result, mathematical techniques originally developed to study tectonic activity have been adapted to study near-repetition phenomena in sociological and behavioral contexts. One of the most important of these methods is the self-exciting point process, also called *Hawkes process*. When there are no prior events (no earthquakes or no crimes), the number of new events to occur per unit time, denoted $\lambda(t)$, is given by a constant $\lambda(t) = \mu$. This indicates that at the beginning of the process, events will occur at any time with the same probability. However, if there is an initial event at time t_1, a first earthquake or a first crime, the number of future events per unit time is temporarily larger and given by

$$\lambda(t) = \mu + \alpha\, e^{-\beta(t-t_1)}.$$

Note that immediately after the first event $\lambda(t_1) = \mu + \alpha > \mu$, and that after a time $1/\beta$ after t_1, $\lambda(t)$ returns to its initial value μ. This means that the first event has acted as a trigger for further events per unit time within the period $1/\beta$. Of course, we do not know the values of α, β, but given a set of geographically clustered events that occur at times $\{t_1, t_2, \ldots, t_i, \ldots\}$ that are part of the same process and affect $\lambda(t)$ in the same way, these events can be fitted to the following form

$$\lambda(t) = \mu + \sum_{t_1 < t} \alpha\, e^{-\beta(t-t_1)},$$

from which α, β can be derived. Large values of α and small values of β mean that the near-repetition effects are large and last longer. One can also estimate the number of „child„ events triggered by a single „parent„ event and show that this is given by $\gamma = \alpha/\beta$. Modeling of self-exciting point processes was applied to residential burglaries, violent civilian deaths in the Middle East, gang violence, terrorist attacks, and protests in the wake of lockdown efforts to combat COVID-19 in various countries.

A final issue is that of rehabilitation. For a long time, especially in the United States, the primary purpose of incarceration was to punish offenders. Harsh sentences were seen as a deterrent to future crime, not only for the offender himself, but also as an example for others to refrain from criminal activity, since they might spend long periods in prison if arrested. However, the results have not been encouraging, as crime rates have increased throughout the United States in recent decades. Perhaps another way is possible? Indeed, rehabilitating offenders may be a more useful way to reduce crime by helping people change their behavior through job training to prepare for jobs outside society, substance abuse counseling, helping people manage anger and violent impulses, and building parenting skills. So the question becomes „stick or carrot“: Is it better to punish offenders (the stick) or help them rehabilitate (the carrot) to reduce crime?

One way to address this question is through evolutionary game theory, which assumes that individuals make decisions, in this case to commit a crime or not, depending on three factors: their prior arrest history – high prior sentences reduce the likelihood of recidivism – the social environment – the more criminals present, the greater the likelihood of committing a crime – and the resources for rehabilitation offered to individuals with prior arrests. Another possibility is that individuals with criminal pasts permanently leave their criminal lives (and the evolutionary game) as reformed individuals. Finally, it is assumed that once an individual has committed a certain number of crimes, he or she will not change his or her behavior and will also leave the game, but as a recidivist. The sociological factors that determine which decisions are made (e.g. the resources for rehabilitation or the amount of punishment) are represented by mathematical parameters. Changing them leads to different outcomes, with different proportions of individuals who never commit crimes, who improve, or who remain incorrigible until the end of the game.

An important mathematical quantity is the final ratio between those who never commit crimes and those who improve, P, the „paladins“ of society, and the recidivists U, the „unreformed“. Large values of P/U represent „good“ societies with many more virtuous people than hardened criminals, and small values of P/U represent „bad“ societies with many criminals and much crime. In an ideal world, one would have all possible resources to adequately punish and rehabilitate criminals. In the real world, however, total resources may be limited, so if society decides to increase the costs associated with punishment (e.g. by lengthening sentences or keeping criminals under high security for long periods of time), the costs associated with rehabilitation will necessarily be reduced.

We find that the most successful strategy for increasing the P/U ratio and reducing the number of criminals (and crime) is to optimally allocate resources so that criminals experience effective intervention programs after they are punished, particularly in the early stages of their return to society. Overly harsh or lenient punishments are less effective. Thus, the dilemma of „stick versus carrot„ can be solved by a „stick and carrot„ strategy: The best way to reduce crime is to ensure that punishment is harsh enough to deter criminals from reoffending. At the same time, it is important to ensure that sufficient resources are available for rehabilitation.

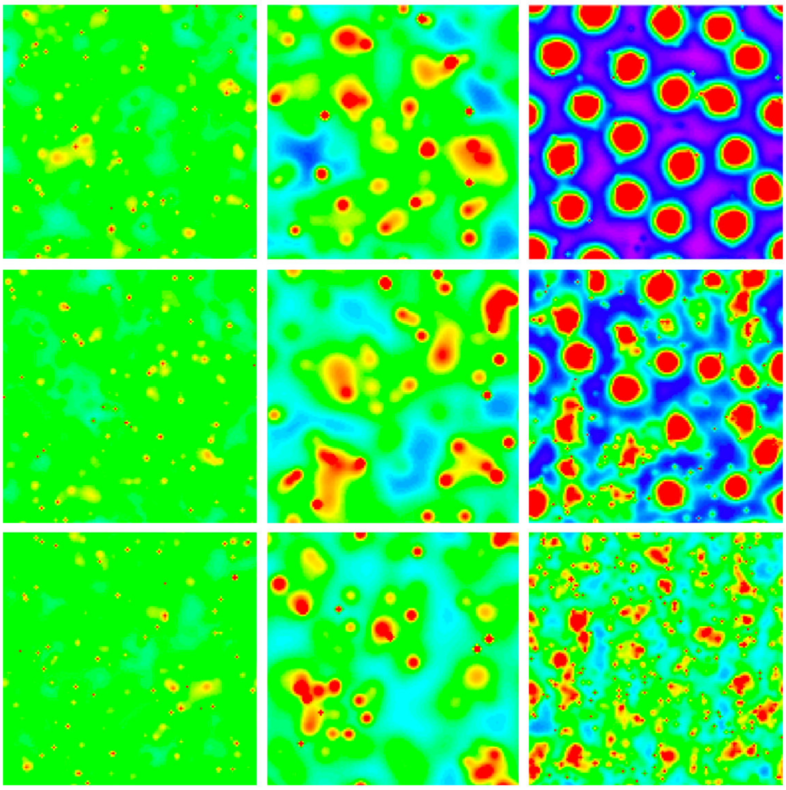


Figure 83: Crime-induced hotspots in the attractiveness field, starting from a uniform value: initially, all locations in the virtual „city" are equally attractive. When there are many criminals but the attractiveness field does not increase significantly after a crime, hotspots never occur (top row); when there are many criminals and the attractiveness field increases significantly after a crime, hotspots occur and are stationary (bottom row). When there are few criminals, hotspots appear only when changes in the attractiveness field are large, but they are transient (middle row). The colors follow the rainbow spectrum from purple (minimum) to red (maximum), with green being the center.

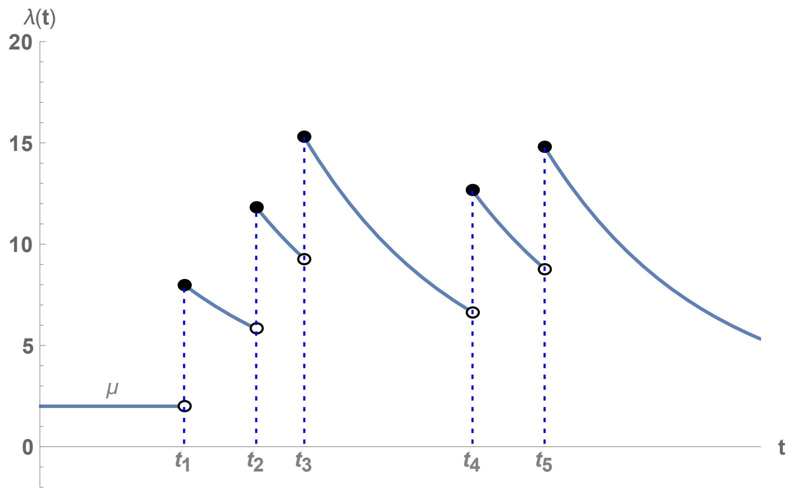


Figure 84: Dynamics of a self-exciting process. Up to the first event (an earthquake or a crime) at time t_1, the number of expected events per unit time is a constant, $\lambda(t) = \mu$. After the first event at time t_1, $\lambda(t \to t_1^+)$ increases to $\mu + \alpha$, as indicated by the first jump at t_1. The magnitude $\lambda(t)$ then begins to decrease, so that if there were no further events, it would return toward the initial value μ after a time $1/\beta$. However, a new event at time t_2 causes a second jump in $\lambda(t)$, so that $\lambda(t \to t_2^+) = \mu + \alpha\, e^{-\beta(t_2 - t_1)} + \alpha$. The process continues in this way. In this picture $\mu = 2$, $\alpha = 6$, $\beta = 0.15$ and the series of events takes place at times $\{t_1, t_2, t_3, t_4, t_5\} = \{6, 9, 11, 18, 21\}$. All times are assumed to be in arbitrary units in this representation.

59 The Offender's Home: what is Geographic Profiling?**

With a serial killer, it's hard to predict who the next victim will be, but arithmetic can tell where the perpetrator lives.

In the very first episode of the crime series *Numb3rs* a serial killer defies the police. The detectives look for a pattern in the places where the bodies are found. In this way, they hope to predict where the killer will strike next time. The brother of a police detective, who happens to be a mathematician, explains that this is nonsense: „If you take a garden sprinkler and look at the pattern of the drops, it is impossible to predict where the next drop will land. But the pattern of drops will tell you where the sprinkler is."

Then, in accordance with this equation, the mathematician creates a model that predicts where the killer will live. The police search the neighborhood in question, but DNA searches show that the killer does not live there. The mathematician adjusts his model to account for the difference between living and working, and in the now specially designated area, the police know where to find the killer. Then it turns out that he has just moved and that the first model had properly designated his old street as suspicious.

The TV producers didn't have to come up with much for this story, because that's almost exactly how it happened in real life. Only in this case, it wasn't a police officer's brother who provided the solution. During his promotion in 1991, criminologist Kim Rossmo developed a formula, to determine where the perpetrator is likely to live from the crime scenes. The formula consists of two terms. One term describes that the offender will not strike too close to his own home. The other term states that after this buffer zone, the probability of committing a crime slowly decreases. Later refinements of the formula also take into account the severity of the crime. The further away from home, the more violent the crime. The result of the formula is exactly as in *Numb3rs*: a colored map that shows per district how likely it is that the offender lives there.

Just like in the Movies

Kim Rossmo is now a professor at Texas State University and directs the *Center for Geospatial Intelligence and Investigation*. The episode of *Numb3rs* looks suspiciously like a search for a rapist, who was active in Louisana for ten years. In 1998, a then rather desperate police officer asked Kim Rossmo for help. Rossmo collected data for a few days, then created color maps showing where the perpetrator probably lived. That left an area of about one square kilometer to investigate: the danger zone. But all the men in the neighborhood turned out to be innocent after DNA testing.

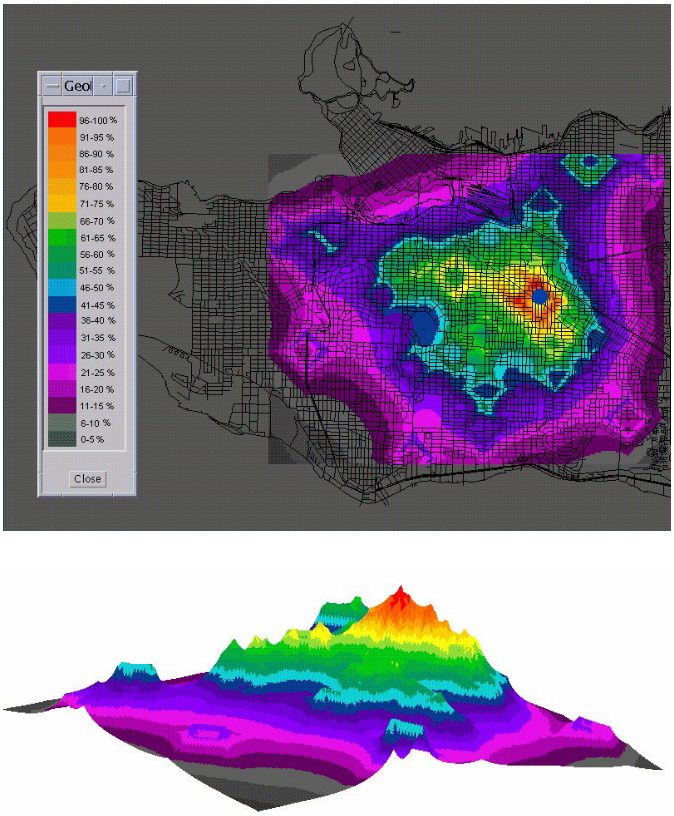


Figure 85: Images of a geoprofile and hazard area for a series of robberies in Vancouver, BC, Canada. The blue dot on the geoprofile indicates where the perpetrators lived.

Then the police received a tip about a new suspect, who, it turned out, did not live in the area Rossmo had designated at all. However, a clever detective discovered that the man had just moved and lived in the middle of the danger zone. Just as in *Numb3rs*. DNA on a discarded cigarette proves that this suspect is indeed the culprit.

The Only Deductive System

Kim Rossmo hopes his work will clarify the science of *Geographic Profiling*. Much forensic science is based on deduction rather than logical reasoning: It is inductive rather than deductive. Deductive reasoning is based purely on facts and logic. Think about it:

> *All human beings are mortal. Socrates is a human being. Socrates, therefore, is mortal*.

Inductive reasoning is based on observations from which you draw the most plausible conclusions possible. Think about this:

> *For the past week, my neighbor let his dog out every morning at eight. So tomorrow morning, the neighbor will walk his dog at 8:00 a.m.*

Inductive reasoning often works well in practice, but it is not watertight. In the example above, it may well be that the neighbor oversleeps. Rossmo: „Most science is induction: you record your observations and make generalizations from them. The only real deductive system is mathematics."

Meanwhile, Rossmo's system has helped solve hundreds of cases. Rossmo emphasizes in interviews that he is not concerned with predictions. On the contrary, his model goes back to the starting point. To explain this, he uses the metaphor of the garden sprinkler. He jokes that he is 99 % sure that the authors of *Numb3rs* got this idea from an interview. The creators really didn't have to come up with much themselves.

60 What are the Chances that the Suspect's DNA Trace will be found?*

Suspects of a crime often leave behind clues ranging from shoe prints to clothing fibers and DNA traces. Forensic statistics are becoming more accurate in calculating the probative value of such traces.

Since 1988, DNA analysis has been used in Germany to prove that criminals have committed a crime. The most common procedure is to compare a DNA trace found at the crime scene with the DNA of crime suspects or with the DNA analysis file (DAD), in which, among other things, the DNA patterns of crime suspects who have already come to light and of other crime scenes are stored.

The DNA that humans carry in almost every cell of their bodies consists of several billion individual parts, which are arranged differently in double strands in each person. There are four basic building blocks, which repeat themselves differently in certain sections. This different number of repetitions in precisely definable areas makes it possible to determine so-called DNA patterns, which are individual to each person, see Figure 86. It may well be that the number of repeats in certain sections is the same in different people. Population studies have determined what the probability is for a random person in the population for a given number of repeats in each section.

Figure 86: DNA-Muster

While the DNA identification pattern was initially limited to five sections, at least twelve comparison sections are now recorded in the EU. In addition, more can now be analyzed from the smallest traces than was previously possible. The more sections that can be processed from a trace found, the more reliably statements can be made as to whether the comparison person also caused the trace. „Even if a DNA trace at the scene of a crime does not necessarily prove the perpetration of the crime, it is still important evidence or a starting point for further investigations“, says the young detective Befeldt.

Investigation of a Murderer in the Netherlands by DNA Analysis

In 1999, Marianne Vaatstra, then a sixteen-year-old girl, was raped and murdered, as she was riding her bicycle home. There was a DNA trace on her body that most likely came from the perpetrator. In the following years, the prosecution arrested at least twelve suspects, but the perpetrator was not found. In September 2012, an investigation began into the DNA supremacy. The idea was that the perpetrator might be tracked down through family members. DNA analysis techniques can provide clues, based on the DNA trace left behind, as to whether the possible perpetrator is genetically related to one or more participants in the DNA population screening. 7,200 men living within five kilometers of the crime scene volunteered their DNA. The Netherlands Forensic Institute (NFI) compared the DNA profiles of these volunteers with the DNA trace of the perpetrator. Unexpectedly, the NFI found a direct match with the DNA of one Jasper S. He had voluntarily participated in the population screening and did not even need to be traced through family members. In December 2012, Jasper S. confessed to the murder. He was sentenced to eighteen years in prison.

Solving a Murder after 25 Years

In Germany, too, so-called cold cases are being solved by DNA analysis. In 2020, for example, the murder of Brigitta J. in Sindelfingen was solved. After the investigation had been discontinued without success in 1995, new comparisons of the DNA traces found on the victim at that time led to the crime suspect Hartmut M., whose DNA traces were available in the DAD, as he had already been convicted of a homicide in 2007, see figure 87. Kriminaldirektor Mathias Bölle, the head of the Böblingen Criminal Investigation Directorate, says of this success: „This case shows us once again what outstanding results the persistent and intensive work of our colleagues in combination with modern forensic methods can lead to.“

Figure 87: Hartmut M.

Blood Group

The clearances of these cases are just two of the many examples of successful application of forensic DNA analysis to, for example, saliva, hair, blood, or semen left behind. „In recent years, forensic research has increasingly evolved from a craft to a science“, says NFI statistician Marjan Sjerps, also a part-time professor of forensic statistics at the University of Amsterdam. Mathematics, especially probability theory and statistics, play an important role in this science. Sjerps: „Thirty years ago, the most accurate statement we could make was that the blood type of a lead matched that of a suspect, and that the probability of a coincidence, for example, was ten percent. With today's DNA profiles, that probability is less than 1 in 1 billion.“ This probability determination is used in a particular type of probability model: *Bayes' probability model*. This model has the important property that the probability of a hypothesis is naturally adjusted as additional evidence becomes available over time.

When a DNA trace is found at the crime scene, the prosecution asks the NFI, analyze the trace and determine how strong the evidence is that some or all of the DNA came from the suspect. The strength of the evidence is measured by what is called the *likelihood ratio (LR)*. The larger the LR, the stronger the evidence that the trace really came from the suspect. The LR is the ratio between two probabilities. The probability in the numerator is the chance that there is a match if the trace comes from the suspect with certainty. This chance is almost 1 if it is a very nice unmixed trace and if the analysis is done perfectly. The probability in the denominator is the chance that there is a match if the trace is not from the suspect with certainty. The rarer the DNA profile, the lower this second chance. Sjerps: „Reliably determining these two probabilities for all types of leads is a major challenge for forensic statistics.“

Identification of Victim

The same type of probability calculation used in the Vaatstra case was successfully used by the NFI to identify the victims of the 2010 plane crash in Tripoli, Libya. This resulted in the deaths of 71 Dutch nationals. Only one person survived the accident, a nine-year-old Dutch boy named Ruben. The NFI used DNA from family members, such as parents, brothers and sisters, to identify the victims. Sjerps says that in the coming years, forensic statistics will become more accurate in determining the probative value of more and more different traces: „Thanks to new technological developments, we will be able to determine the DNA or chemical composition of smaller and smaller traces, for example. We will also be able to analyze mixed traces – such as a cigarette shared by two smokers – with increasing accuracy. In addition, NFI now produces separate reports on the probative value of fingerprints, for example, DNA traces and clothing fibers from the same crime scene. In the future, we will also calculate the combined probative value of these traces.“

61 How does Shazam recognize a song so quickly?*

Every song has a unique digital fingerprint. Shazam knows the formula for this fingerprint and uses it to recognize songs.

I hear a beautiful song echoing from the speakers and point my phone's microphone at it. I press 'Shazam', see the app pick up the sound and even respond to the different pitches, and after a few seconds, „The Heart Of Saturday Night – Tom Waits“ appears on the screen. Even the album cover that contains the song appears on the screen. It's a magical moment every time Shazam tells you who's singing the song and what the title is.

Shazam is an original American company founded in 1999. Since 2008, the company has offered its music lovers a free app for cell phones. Shazam now has a database of more than eleven million songs whose digital fingerprints have been identified. Only songs from this constantly growing file can be recognized by Shazam. Every day, Shazam is accessed up to fifteen million times worldwide. Even the background noise in a café or a bit of sound distortion does not confuse the music lover. What is the secret?

Digital Fingerprint

The secret of Shazam is the algorithm it uses to create a digital fingerprint of a song or piece of music. This digital fingerprint is a unique sound pattern of a song, just like a fingerprint is a unique line pattern on each person's fingers.

You can create a spectrogram from any sound recording: a three-dimensional graph in which on the x-axis is the time, on the y-axis the frequency of the sound, and on the z-axis the strength of the sound. So each point on the graph tells what the frequency and amplitude of the sound is at what time. Although this spectrogram contains all the characteristics of the sound, it contains so much information, that it is inconvenient to store it all digitally for later reference.

Shazam therefore focuses on the intense fragments from a song, say the peaks in the spectrogram. The algorithm determines the time of these peaks and the frequencies associated with them. The result is a two-dimensional graph of points that stand out in the spectrogram, as if you were just displaying the peaks of a mountain landscape on a map. This information is usually not enough to identify a song. That's why Shazam's designers have used a clever trick. They calculate how each peak relates to other nearby peaks. This is determined by the time between peaks and the frequency differences between them.

The information about individual peaks and their relationship to nearby peaks determines the digital fingerprint, which turns out to be (almost) unique in practice. Shazam compares the digital fingerprint of a song you record as a user, with all the fingerprints available in its database. If there is a match, the title and artist appear on the screen.

Hit Prediction

Although Shazam works well for pop music, the music recognizer has trouble with classical music. Shazam on my smartphone recognizes a piece from Mahler's first symphony, but when I show it a piece from Sibelius's first symphony - also a powerful, distinctive piece - it has no idea. That's not surprising, because many pieces of classical music don't last a few minutes like pop music, but sometimes an hour. If you hear only ten seconds from an hour-long piece of music, it's much harder to determine which piece it's from. Also, how a piece is performed makes a big difference. It can vary greatly from orchestra to orchestra and conductor to conductor.

One of Shazam's newest applications is predicting hits. To do this, the company combines online song reviews with how much Shazam is used for a particular song: When listeners upload a piece of a song to Shazam to know who is singing it, they usually do so because they think it's a good song. And if the number of people who want to recognize the same song increases dramatically within a few days, the song could become a hit.

62 How Fair is the Singing Competition?**

At the Eurovision Song Contest, the discussion about nepotism flares up again every year. An analysis shows how fair the awarding of points really is.

Every year dozens of countries send a song to the Eurovision Song Contest (ESC). The countries give each other points and at the end of the evening there is a winner. Afterwards, people often grumble about these points, especially if their own country didn't win. That's not fair: the Eastern Bloc countries only vote for each other, just like the Scandinavians. It's all nepotism.

Greece and Cyprus are also notorious for always giving each other lots of points. In 2003, for example, Greece ranked 17th out of 26 with 'Never let you go' (who remembers?). The ballad received a total of twenty-five points, but Cyprus, as every year, gave the maximum of twelve points to its neighbor. Surely this must be a set-up?

What is Fair?

Mathematician Michel Vellekoop and econometrician Laura Spierdijk discovered at the coffee machine at the University of Twente that they were both fanatically following the song festival. They got to talking about the so-called unfair distribution of points. They wondered if the complaints about this were true. They decided to find out in a clean, scientific way whether the allocation of points was fair[61].

One problem was that the quality of a submitted song was unknown. Vellekoop and Spierdijk solved this problem by using the average score of a song as a measure of quality. Then they looked at how individual countries deviated from that average, in other words, how strong their preferences were. Vellekoop: „The Netherlands gives an average of 2.5 points too much to Belgium and 1.6 points too little to Malta. Of course,

[61] L. Spierdijk, M. Vellekoop, *The structure of bias in peer voting systems: lessons from the Eurovision Song Contest, Empirical Economics*, 36(2) (2009), 403-425.

it's nice to see who gives each other a lot of points, but to say anything about the underlying structure, you need to do more and formulate a model."

Many existing articles look only at aggregate data. But when you look at the statistics of all countries together, you sometimes see things that can be explained by other variables as well. For example, two neighboring countries may give each other a lot of points because they have the same taste in music. So Vellekoop and Spierdijk decided to look at the different countries separately and determine which variables best explained their preferences. „What we had to mention came quite directly from folklore. Do neighboring countries give each other more points? Does religion matter? How important is language?"

Defining these variables precisely proved to be a lot of work. How do you measure how close two countries are? And are countries with a sea between them considered neighbors? If Britain is a neighbor of the Netherlands, then it is also a neighbor of Italy. The sea neighbors eventually fell into the model and for the distance between countries, the distance between their capitals was taken. It was also funny that there were already many definitions for other things. For example, it turned out that there were measurement systems that indicated how much two languages resembled each other.

A Meaningful Theory

Vellekoop was surprised when he discovered how they could combine all this data: „I always felt that you had to be careful with dependencies and that it was a problem that we used all kinds of variances and scales interchangeably. But then it turned out that there is a theorem that says that by making a few very weak claims, you can still approximate the good parameter and make a fairly consistent estimate of all these variances. In econometrics, this is all very well known, but as a mathematician, it is very nice to see how powerful the theory turns out to be in a different area."

Their model showed which variables have a big impact in each country. For example, the Netherlands gives a lot of points to a solo singer who sings in French. Vellekoop does not call preference for one language unfair: „You can't help it if you find French more beautiful." Incidentally, this language preference proved to be the strongest in Greece and Cyprus. They punish countries that sing in a language that is far from their own. The fact that these countries give each other so many points is not just because they are neighbors. Language is a stronger explanatory variable.

In general, neighbors seem to give each other slightly more points, but the variable „neighboring country" did not stand out enormously in most countries. In the Scandinavian countries, mutual points seem to be better explained by similar musical tastes than by nepotism. Major exceptions are Estonia, Latvia, and Lithuania. Vellekoop: „They love their neighbors very much: On average, they get almost four points too many. Our overall conclusion was that the Baltic states have something to explain about their voting behavior."

Analysis of Social Networks

The 2010 U.S. film drama „The Social Network“ describes the genesis of the social network Facebook. The analysis of social networks examines Connectedness of group members with certain commonalities. The links include kinship relations, informational connections, or physical approaches. The goal here is the identification of the leader of the network or the role of subgroups. role of subgroups (so-called clusters) in the network. In reality, these networks could be gangster groups, co-authorship (Citation index)[62], online networks (Instagram, Facebook), email communication, pass structure of a soccer team[63], etc.

To analyze networks mathematically, one uses *graph theory* and *centrality measures*. A graph consists of a set of points (so called *nodes*), usually persons. Each two nodes are connected (or not) by an *edge*. The edges show the flow of information within the group. Here, for the singing competition ESC, we consider directed, simple, complete graphs. One needs special quantities to be able to evaluate the people in the social network.

Centrality Measures

The analysis of a network is based on centrality measures; they say something about the 'importance' of actors in the network, such as the extent to which actors concentrate relationships on themselves or the possibility for actors to disrupt the relationships of others by their location in the network. From the centrality measures of actors, one can infer the distribution of power in the network. The most important three measures are.

- Degree centrality: direct connections in the network.
- Closeness centrality measures how fast one node can reach another
- Betweenness centrality: What potential does an actor have to control the flow of interaction within the network?

Degree Centrality

Degree centrality is the sum of an actor's direct connections. This number of direct connections an actor has with other is considered a measure of an actor's activity in the network: the more relationships an actor has, the more central it is. However, direct connections alone are not decisive. A path is a path over multiple edges connecting 2 vertices and the path length is the number of edges it contains. The distance between A and B, d(A,B), is the length of a shortest possible path from A to B; such a path is called *geodesic*.

[62] A.L. Barabâsi, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, T. Vicsek, *Evolution of the social network of scientific collaborations*, Physica A: Statistical Mechanics and its Applications, 311(3-4) (2002), 590-614.

[63] C. Cotta, A.M. Mora, J.J. Merelo, C. Merelo-Molina, *A network analysis of the 2010 FIFA world cup champion team play*, Journal of Systems Science and Complexity, 26(1) (2013), 21-42.

Closeness Centrality

Closeness centrality is 'autonomy in the network'; it measures the speed of interaction in the network. Here we include indirect connections via third parties are included. Closeness centrality describes the speed with which it is possible for an actor to reach any other actor: the fewer detours it has to make via third parties, the more central the actor is. The closer a point is to all the others, the more effectively and and more independent it can move in the network; it is less dependent on the reliant on the willingness of others to cooperate. This *Closeness score* of node C is calculated as C=1/d(C,A) + 1/d(C,B)+ ..., i.e. the smaller the distance to C, the larger the reciprocal. Here, not only immediately adjacent nodes are evaluated.

Betweenness Centrality

Central are the actors through which the most connections run, and this is to be measured with betweenness centrality. What is important here are the indirect relationships through third parties that an actor has on him. Similarly, the ability to control the flow of interaction of an actor as a third party involved is the focus. Actors are the more powerful, the more shortest connection paths they can interrupt between other points. The betweenness score of C as a link between A and B is the number of geodesic paths from A to B that pass through C divided by the number of all geodesic paths from A to B. Then the total betweenness score of C is the sum of of all individual calculations of all possible nodes A, B.

Eurovision Song Contest (ESC)

The ESC is a network of countries with geographic/historical relationships that exchange scores among themselves. One can use a dynamic network to: analyze voting behavior in the Song Contest over several years. However, there are two problems: constant rule changes and changing finalists (except for the 'Big Five': Germany, France, Spain, UK and Italy).

In the basic scoring system (since 1975) each country gives points to 10 other countries, for example. Country A gives points $\{1, 2, 3, 4, 5, 6, 7, 8, 10, 12\}$ to the 10 (best?) countries $\{B, C, D, E, F, G, H, I, J, K\}$. Countries are not allowed to choose themselves. The representation is done by directed, weighted, dynamic graphs, namely with directed edges for the point assignment and weighted by the respective score.

clique formation at the Eurovision Song Contest

Usually, in social networks, two nodes with a common third node have a higher probability of being connected. If there are cliques in ESC, the observed clustering coefficient will be higher than the coefficient from a 'random contest'. If two countries both vote for a third country, then they are more likely to vote for each other. The so-called *clustering coefficient* C of a node ν is the probability, that two neighbors of a given node ν

are neighbors themselves. To determine C, one counts edges between neighbors of a given node ν and divides this by d(ν)!/2!(d(ν)-2)!. The exclamation point here denotes the factorial. Averaging C over all nodes gives the clustering coefficient of the network.

Analysis of the Eurovision Song Contest 2010

Finally, we want to analyze the ESC 2010. The question is whether Lena Meyer-Landrut won by clique formation. For this we used the free software SovNetV, version 2.9 and visualized the whole voting behavior in a graph, see Figure 88.

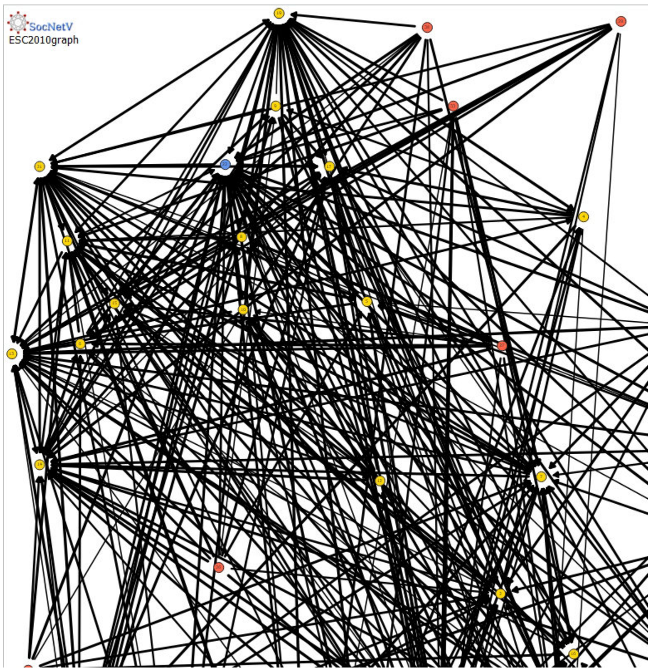


Figure 88: ESC 2010: directed graph modeling voting behavior.

Germany is the blue node with the number 22. The red nodes gave points to Germany, the yellow ones got and/or gave points. One can now with the software SocNetV the many centrality measures, e.g. „P „Prominence/Betweenness Centrality“ or „Communities/Cliques“, see Figure 89. For the latter, Lena has a clique size of 34, which is an unremarkable value here: she thus won the ESC fairly! By the way, for experimentation, the used data file ESC2010 for SocNetV is provided on the book’s web page.

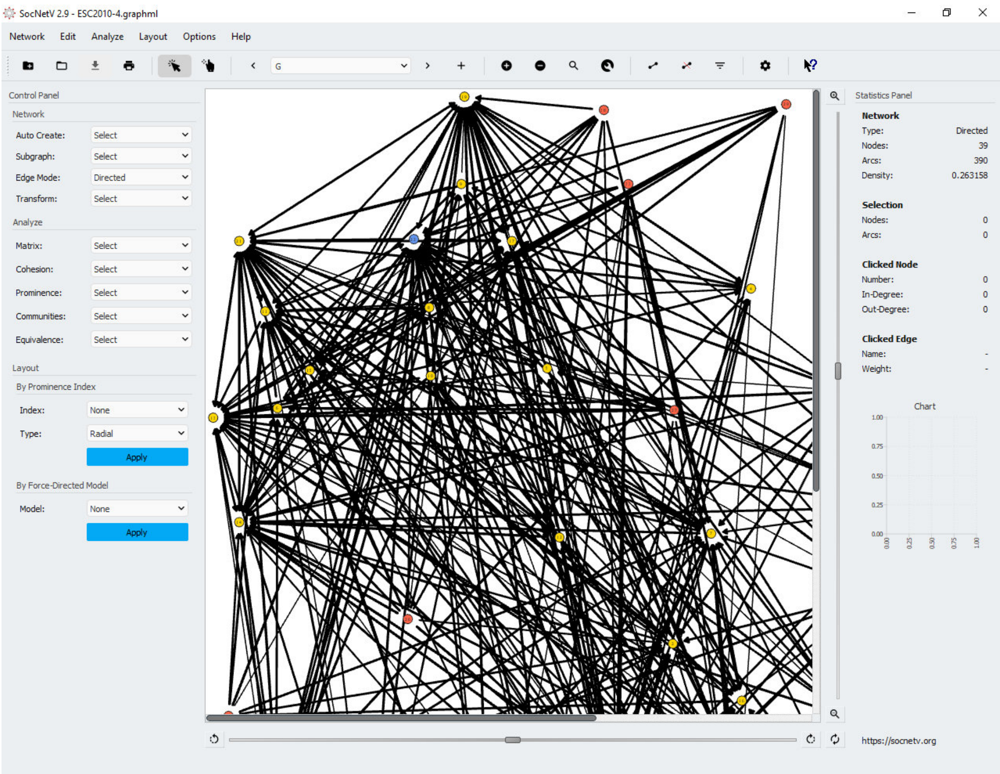


Figure 89: SocNetV software with the ESC 2010 graph.

63 Origami Mathematics at the Museum of Modern Art*

The mathematics behind the folding of origami patterns not only inspires beautiful art, but also leads to a better understanding of the folding of proteins, the workhorses in the human body.

Origami is an ancient Japanese art form based on folding paper. Origami mathematics has only been around for about forty years. Origami mathematics describes the rules of origami in a formal way and discovers what is possible and what is not. With origami mathematics, mathematicians try to understand how to fold any three-dimensional object from a flat piece of paper: a frog, a crane, a human being, or an abstract geometric figure. It can be anything.

The ultimate challenge is to find a method of calculation, that tells you how best to fold any three-dimensional structure. And „am best" means something like with as few folds as possible and as small a piece of paper as possible.

Origami Loops

American mathematician Erik Demaine (1981) uses mathematics to fold beautiful new shapes out of paper. Together with his father Martin, who is an artist, he designed and folded three origami shapes - called „Computational Origami" which can be admired in the permanent collection of the Museum of Modern Art in New York. Each form connects several circular pieces of paper to form a fascinating *looping*. Follow the *looping* imaginatively with your index finger, and you will find that the total angle of rotation is much greater than 360°, usually something between 720° and 1080°.

This form of origami differs from traditional origami in that the shapes are actually made up of an infinite number of infinitesimal folds. This is the only way to create a surface that curves everywhere. Erik Demaine: „The underlying mathematical question is how mathematics can provide a computational recipe that specifies, which curved shapes can theoretically be folded.“

Origami math doesn’t just help artists. It is applied in all fields related to folding: For example, in robotics, but also in architecture, fine arts and computer graphics. For example, origami artist Robert Lang is working on a foldable telescope lens for use in space. There, one must be able to fold the lens to a plane a hundred meters in diameter, but folded up in a spaceship, you want to keep a package no more than ten meters wide. How do you do that in the most convenient way?

Figure 90: „Computational Origami“ by Erik and Martin Demaine. These three origami figures are part of the permanent collection of the Museum of Modern Art in New York. Source: Erik and Martin Demaine.

Folded Proteins in Your Body

Another important practical problem Demaine is working on with origami math is the problem of protein folding. Proteins are the workhorses of the body. They are complex, three-dimensional molecules, that are folded in a particular way. To a large extent, this folding determines their function. In all kinds of diseases, something goes wrong with protein folding, and drugs can fix that with specially designed proteins.

As an origami mathematician, Demaine poses the question, What is the best way to fold one protein shape into another? „Mathematically, this is a kind of one-dimensional origami, because you can think of a long molecule as a line. The question is whether we can find a computational method for the most convenient way to fold the protein. We hope that when we find this mathematical method of calculation, we will have automatically gain insight into the natural and chemical principles that determine how a protein folds.“

Father and son Demaine experiment a lot with paper to solve a mathematical origami problem. Son Erik: „Folding an origami object with his hands forms the intuition for this, what is possible and what is not. It’s like building a big experimental database by folding it in your head. You use this database to solve the mathematical puzzle. By the way, all the famous origami designers use some form of origami math. Only they don’t let a computer do the math, they do the math unconsciously in their heads, in an intuitive way.“

64 Love Poetry: Petrarch and Laura – Chaos in Love Affairs***

Dynamic systems describe love affairs to help sort poems from the Middle Ages.

In this chapter we want to discuss to what extent one can love can be regarded as a (deterministic) mathematical model. Although love is often described as something spontaneously controlled, without any structure („love at first sight“), we want to set up a model and calibrate it using famous examples of relationships like Romeo and Juliet, Petrarch and Laura or Jack and Rose (movie 'Titanic'). In doing so, feelings such as fear, affection, longing, desire and love is defined as a reaction to these feelings. defined.

Steven Strogatz is considered one of the pioneers in this field: he formulated a system of differential equations to describe the dynamics of love. What was initially intended as an amusing motivation for his students quickly developed into its own field of so-called dynamical systems[64].

The Basic Model

Taking into account personalities, Strogatz's model describes the interaction of two people, and in the following 2 variables, the lovers. One's own personality plays a role in the reaction to the feelings of the other person; for example, a narcissistically inclined person cannot reciprocate the feelings of the partner. Nevertheless, he loves the feeling of being loved himself. A self-conscious person would always respond to advances from the beloved with advances.

One's own life history can also influence these reactions influence; if one was disappointed or offended by one's first love, one is a bit more cautious in the next relationship and does not overreact to the other person's feelings.

In this model, timely reactions are assumed; for example, no long-distance relationships can be simulated, since there is a loss of feelings over time. Love in the model is the direct reaction to an action of the other person, which leads again to a reaction and to an action etc.

It is also possible to simulate feeling states that take time to arise or disappear, e.g. depression or emotional numbness. The most important assumption is that the simulated persons can read and interpret emotions. Otherwise, a reaction would not be possible and the model would be meaningless. We mention that the model does not take into account the general gut feeling, but only the external reactions.

The two following models are linear, that is rather simple, so that further analysis is still

[64] S.H. Strogatz, *Love affairs and differential equations*, Mathematics Magazine, 61(1) (1988), 35-35

possible. They have the form

$$\dot{X}(t) = \alpha_1 X(t) + \beta_1 Y(t),$$
$$\dot{Y}(t) = \beta_2 X(t) + \alpha_2 Y(t).$$

Here $X = X(t)$ and $Y = Y(t)$ describe the feelings of the two loving persons: Love when positive and Dislike when negative. The factors α_1, α_2 describe the influence of the own and β_1, β_2 the influence of the feelings of the other on the current love. The dot above X, Y represents the time derivative.

Romeo and Juliet

This model was used by Strogatz to motivate his students. It describes in a very simplified way the course of the love relationship of Romeo (X) and Juliet (Y) after the play by William Shakespeare: Juliet is courted by Romeo, however their families are enemies and forbid this love. Finally, both elope and the tragic death of both occurs. One could therefore take as a parameter:

$$\alpha_1 = 0.05, \quad \alpha_2 = -0.3, \quad \beta_1 = \beta_2 = 0.1, \quad X(0) = 9, \quad Y(0) = 3.$$

The positive β_1, β_2 describe the strong reaction to the feelings of the other. Romeo is a young man who is only spurred on by the feelings of both, so he has a positive α_1. In contrast, Juliet has a negative α_2, thus reacts negatively to her own feelings, because as a young girl she does not want to disappoint her family and thus does not accept her own feelings. The initial values $X(0)$, $Y(0)$ were chosen considering physical attractiveness. The time interval is based on the play 30 days.

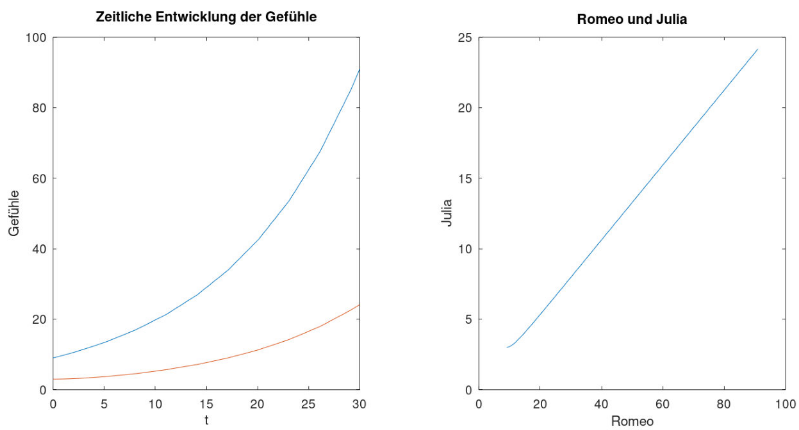


Figure 91: Temporal evolution of emotions (left) and Romeo vs. Juliet (right).

You can see in Figure 91 that with simple modeling, love develops in a very linear way. It rises steadily after their first meeting, and the two lovers reinforce each other. There are no ups and downs in their mutual affection, no wavering of either party, or other real interaction as occurs in Shakepear's play.

To have a closer love progression to the play, the model can be extended as follows:

$$\dot{X}(t) = \alpha_1 X(t) + \beta_1 Y(t) + r_1 A_2 - bY(t-\tau),$$
$$\dot{Y}(t) = \beta_2 X(t) + \alpha_2 Y(t) + r_2 A_1.$$

Here A_1, A_2 denote constants for the attractiveness of the respective partner. The factors r_1, r_2 describe the respective reaction to the attractiveness of the other. The term $-bY(t-\tau)$ is a delay term; it describes a delayed reaction of Romeo to Juliet's feelings ('overthinking').

From the play we infer: Juliet's attraction is higher than Romeo's ($A_2 > A_1$), Romeo's reconsidered reaction to Juliet's feelings is smaller than the direct reaction ($b < \beta_1$). The model also simulates the condition if Romeo and Juliet had not died. They would certainly have separated after a relatively short time, and that would not have been a 'nice' ending for one of the most famous love tragedies.

Petrarch and Laura

The second model was developed by Sergio Rinaldi,[65]. The motivation here was to use a love model to temporally order the disordered poems of the famous medieval poet Francesco Petrarch. These poems result from the inspiration of unrequited love for Laura.

Let us first describe the two characters. Petrarch was a famous Italian poet of the 14th century. He was 23 years old and a not very attractive man when he first met Laura. Like many poets at the time, he was not highly regarded socially. In contrast, 16-year-old Laura was a very attractive, young-married woman and had a high social standing.

Their first meeting resulted in a 27-year unrequited love of Petrarch for Laura, which provided literary inspiration for Petrarch and he began writing love poems. Despite Laura's death after 27 years, Petrarch continued to be inspired for several years beyond that and wrote a total of about 200 poems to her.

Rinaldi's resulting model is based on the basic model with a third equation added:

$$\dot{L}(t) = -\alpha_1 X(t) + \beta_1 A_P + R_L\big(P(t)\big),$$
$$\dot{P}(t) = -\alpha_2 P(t) + \beta_2 \frac{A_L}{1+\delta Z(t)} + \beta_2 L(t),$$
$$\dot{Z}(t) = -\alpha_3 Z(t) + \beta_3 P(t).$$

Here L and P stand for the love of Petrarch and Laura. The factors α_i, β_i, A_P, A_L are defined analogously to those from the Romeo-Juliet model. Here α_i is now the forgetting

[65] S. Rinaldi, *Laura and Petrach: An intiguing case cyclical love dynamics*, SIAM J. Appl. Math. 58(4) (1998), 1205-221.

factor. This means that love fades away after some time without refreshment. Further β_1 stands this time for the reaction to Petrarca's attractiveness and β_2 for the reaction to both the attractiveness and the inspiration and love of Laura. The third equation describes Petrarca's inspiration (Z) and arises linearly from both Laura's and Petrarca's love factors.

The reaction function R_L in the first equation, that is, Laura's reaction to Petrarch's love, can be described non-linearly. It can be derived from the following context: On the one hand, she is flattered by his flirtations. But, as it was usual at that time and class, a married woman had to take these flatteries negatively at some point. On the other hand, when Petrarch's love turns to despair, she eventually feels pity. This pity causes her to respond positively to his flatteries. The resulting reaction function looks like this

$$R_L(P) = \beta_1 \, P\Big(1 - (\frac{P}{\delta})^2\Big)$$

The second equation has been further extended to include inspiration (Z) compared to the basic model. To adjust the constants, it is important to understand the dated poems. From the interpretation of Rinaldi the following relation for the forgetting factor α_i results. It is three times higher for Laura than for Petrarch ($\alpha_1 = 3 \cdot \alpha_2$). Meanwhile, inspiration lasts ten times as long ($\alpha_2 = 10 \cdot \alpha_3$). The love response factors β_i can be described in a similar way. Petrarch's reaction to Laura's love is five times as strong as hers ($\beta_2 = 5 \cdot \beta_1$). The inspiration is even ten times as strong ($\beta_3 = 10 \cdot \beta_1$). This results from the one-sided love. This results in the initial values $L(0) = P(0) = Z(0) = 0$ and the following parameters.

$$\alpha_1 = 3, \quad \alpha_2 = 1, \quad \alpha_3 = 0.1, \quad \beta_1 = 1, \quad \beta_2 = 5, \quad \beta_3 = 10,$$
$$A_P = 1, \quad A_L = 2, \quad \delta = 1.$$

The total duration of the model is the lifetime of Laura, i.e. 27 years. The values were chosen so that the course of Petrarch's love is analogous to the analysis of the dated poems. It is clear that the model captures the relationship well from Petrarch's point of view. Furthermore, it can be seen that Petrarch and Laura are in a whirl of emotions. As can be seen in Figure 92, the love of both is caught in a cycle (so-called LP-circle). Petrarch's emotional world fluctuates from absolute love to absolute despair and back again to a love high. Petrarch's inspiration develops accordingly. One can still adjust the parameters of the model after sorting by the contents of the poems (see Figure 93).

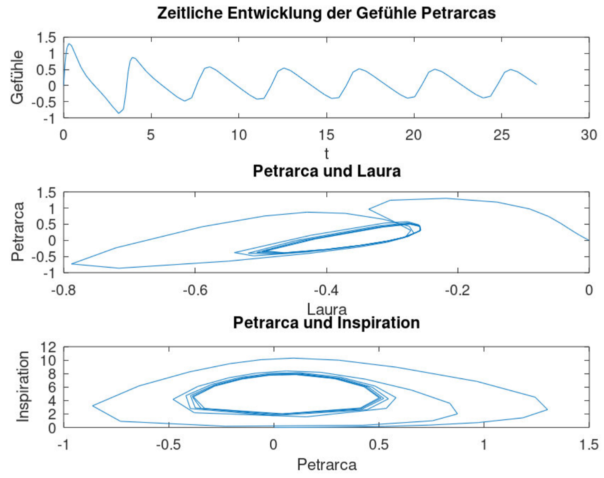


Figure 92: Temporal development of Petrarch's and Laura's feelings.

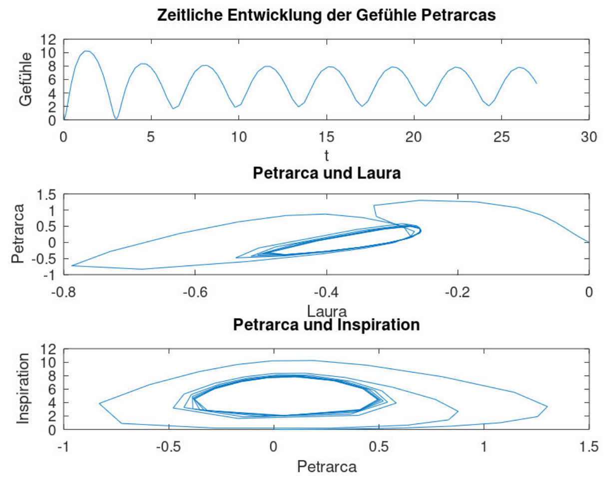
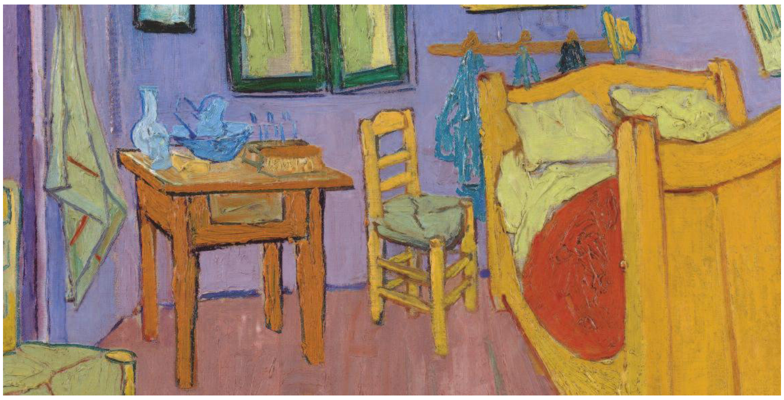


Figure 93: Feelings of Petrarch and Laura after Fit to poems.

65 The Ravages of Time*

Vincent van Gogh's paintings have slowly discolored over the years. Digital reconstructions show how they likely looked originally and how they will fade.



In 1888, Vincent van Gogh described the painting he was working on in a letter: „This time it is only my bedroom. But here the color must be enough. [...] Anyway, seeing this painting must give rest to the head, or rather, to the imagination. The walls are pale purple. On the floor there are red tiles. The wood of the bed and chairs is yellow like fresh butter."[66]

But now, when you look at the painting 'The Bedroom' in Amsterdam's Van Gogh Museum, you see blue walls instead of purple ones. Ella Hendriks, senior conservator at that museum and associate professor of conservation and restoration at the University of Amsterdam, explains that some of the pigments have disappeared. „The paint Van Gogh bought was not always of good quality. For example, he used Geranium Red, a very short-lived synthetic red dye."

Gathering Information

Hendriks was responsible for the restoration of the bedroom and used an arsenal of techniques, to find out the original colors. „We had Van Gogh's letters, of course, and old reproductions also provide information. We also knew that the tape under the frame had been applied during an earlier restoration in 1931. So the colors underneath were what the painting looked like at the time."

Research into the colors of the past was highly multidisciplinary. Microscopes were used to search for pigment grains in the depths of the paint layer. Scans showed which elements were in the paint layers. Hendriks: „Lake Geranium's paint contained the element bromine, and where we found these particles used to be this red color." In the

[66] Letter 705 to his brother Theo (Artes, October 16, 1888) Vincent van Gogh: The Letters, ed. Leo Jansen, Hans Luijten and Nienke Bakker. For the website edition, see: www.vangoghletters.org.

Figure 94: In the center (bottom) 'The Bedroom' as it looks today. On the left (top), the reconstruction of what the canvas originally looked like. On the right (top), the prediction of how the canvas might look in the future after exposure to a certain amount of light. The reconstructions were done by the Van Gogh Museum (Vincent van Gogh Foundation), Amsterdam, in collaboration with RCE (Cultural Heritage Agency), Amsterdam, and Professor R.S. Berns, Munsell Color Science Laboratory, Rochester, NY.

laboratory, the composition of the color of the 19th century was imitated as best as possible and artificially aged.

This gave Hendriks and her team a lot of information, but not yet a complete picture of what the canvas looked like. And that's what the math was all about. This part was done in collaboration with American color scientist Roy Berns. He used the Kubelka-Munk method, named after two German scientists, who created the first theoretical model for the optics of paint layers in 1931 - coincidentally, their work is as old as the earlier restoration of 'The Bedroom'. They use differential equations to describe how paint layers of a certain thickness scatter and absorb light. In other words, how a painting composed of different colors looks.

Past and Future

This is how the reconstruction of the painting came about, how it must have looked when Van Gogh painted it. And then the walls were actually not blue, but turned purple. Hendriks: „Van Gogh is often called a painter of blues and yellows, but in reality he used mostly purples and yellows. That's why this reconstruction is so important for art historians. I now understand much more about where Van Gogh was going, otherwise I wouldn't have known."

By the way, when the canvas was restored, there was no attempt to to restore the colors to their original state. Hendriks: „I'm not going to cover Van Gogh's brushstrokes with my paint. You want to preserve the authenticity of his work as much as possible. Beyond that, some of it remains pure conjecture: We never know exactly what it was like when he painted it." The museum therefore decided to show visitors how the colors have changed over the years in a digital reconstruction.

The advantage of such a digital reconstruction is that it can be extended into the future. What happens if the painting is seen for several hours a day at a certain light intensity? Gradually, the floor gets cooler and cooler, and the red lines in the bed disappear altogether. Hendriks made it a quiz for the museum management: „How much damage do we want to accept? And how long can it take for the canvas to look like this? As conservators, we already knew what was going on, but now we can visualize the damage. These images make much more of an impression than words. The museum is now showing the paintings at an even lower light level, so the discoloration is slowed down."

You can do much more with mathematics, explains Prof. Dr. Ingrid Daubechies (Duke University) in her talk „The Master's Hand: Can Image Analysis Detect the Hand of the Master?"[67] By means of so-called *wavelets*, a mathematical tool used for the analysis and compression of images (also for digital cinema), it is possible to decide whether a painting is an original or whether several parts of a painting were painted by the same person - or not.

In addition, mathematics helps to make overpainted pictures visible. These are first made visible with high-energy X-rays. However, this creates large areas of interference

[67] https://www.youtube.com/watch?v=-SyeN1hCYyA

on the reconstructed image. These interferences can be removed by the mathematical method called *minimization of total variation*.

66 How to Win More Often in Beach Volleyball**

The selection of a promising match plan plays an increasingly important role in elite sports. Mathematical optimization and game theory can provide sound support for this selection.

Dr. Susanne Hoffmeister, Prof. Dr. Jörg Rambau and Ronan Richter (M.Sc.) from the Chair of Business Mathematics at the University of Bayreuth report on research conducted by the chair on sports strategy optimization in beach volleyball. This was the subject of the dissertation project of Susanne Hoffmeister (meanwhile working as Operations Research Scientist at InVision AG in Leipzig), in which optimal game strategies for a beach volleyball final were determined for the first time depending on the specific final pairing.

Olympic final in beach volleyball. We'll be facing the opposing team for the first time in this tournament. Should we take full risk in all rallies, or can we also hope for one or two unforced errors from the others? Should both in our team play equally aggressive?

Figure 95: Germany against Brazil in the beach volleyball final of the Olympic Games, London 2012. Photo by Duncan Rawlinson (https://duncan.co/), without further modifications, License (CC BY-NC 2.0) https://creativecommons.org/licenses/by-nc/2.0/deed.de

Whether you win at a sports game (soccer, tennis, darts, . . .) depends on your own skills, the skills of the opposing team, the luck of the game, but also on the quality of strategic and tactical decisions. According to the 2012 Olympic beach volleyball champions in London, Julius Brink and Jonas Reckermann, with whom we spoke immediately before writing this post, a match plan tailored to the opposing team is a key component to increasing the chances of winning. *Sports Game Strategy Optimization* studies methods of harnessing data from past games to concoct an informed match plan. Such methods have been increasingly used as a supporting tool in competitive sports over the past decade.

In Search of a Success Strategy

Our goal for this post is to explain what kind of information could help one make a strategic decision before a match, how it could be calculated in principle, and how it can be used to determine a strategy for success.

A useful piece of information can look like the color pattern in Table 1. We suggestively call this color pattern in our paper our *winning schedule*.

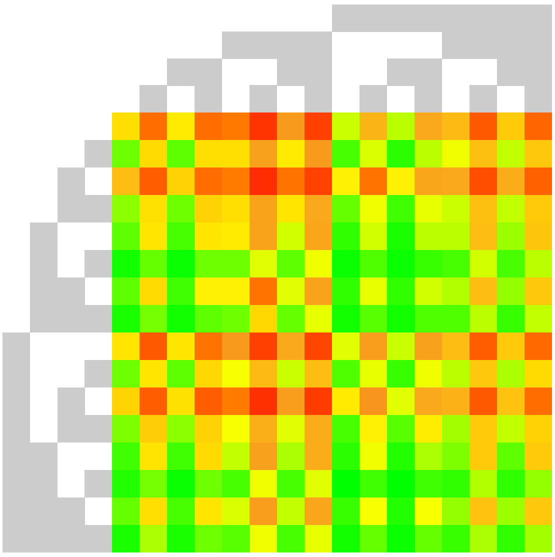


Tabelle 1: Probabilities of winning as a Depending on strategy combinations (rows: our team; columns: opposing team). From green („we win for sure“) to yellow („fifty-fifty“) to red („we lose for sure“).

We can interpret the winning schedule very roughly as follows: The rows belong to our strategies; the columns belong to the opponent's strategies. A color field in the winning schedule belongs to a *strategy combination* of a strategy for us and an opponent's strategy. (We'll resolve later what the gray and white patterns that mark the rows and the columns mean). If we know in advance what strategy the opposing team will play, then we pick the strategy to the row that is greenest in the column of the opposing strategy. Exactly then we have found the success strategy, that is, the strategy with maximum possible probability of winning, against the known opponent's strategy. This success strategy is also called a „best response“ to the opponent's strategy. If we do not know the opponent's strategy beforehand, then we might aim to still do best in the „worst“ case. That is, we choose the strategy to the line where the least green square is still the greenest. Such a winning schedule is called in mathematics (instead of colors there could be numbers, of course) a *two-person zero-sum game in strategic form*. If you have the winning schedule, you have valuable information for a match plan.

But where exactly does the concrete profit schedule in table 1 come from? In a dissertation project at the Chair of Business Mathematics at the University of Bayreuth it was investigated how such a winning schedule for a beach volleyball final can be calculated from observations in the pre-final matches. Susanne Hoffmeister, herself active in handball and triathlon sports at a high level, systematically addressed this question in her dissertation. Table 1 came out as an example. It is the winning schedule for the team Brink/Reckermann (Germany) against Alison/Emanuel (Brazil) before the 2012 Olympic final in London.

So how can we describe a sports game using mathematical methods? Let's take the example of beach volleyball. Beach volleyball belongs to the *rebound ball sports*. This means that two teams facing each other on different sides of the net take *alternating* possession of the ball. (This is also true, for example,in tennis, but not in soccer). According to the rules, the team in possession of the ball may play an attack with a maximum of three alternating ball contacts. At the end of a successful attack, the ball is in the field of the opposing team, which must receive the ball before it touches the ground and can then play an attack in its turn. An unsuccessful attack (usually ball out or ball in the net) results in a point for the opposing team. We call this a „direct point loss“. An attack is especially successful when the opposing team can't handle receiving the ball. We call this a „direct point win“. If the rally continues, we name that a „ball in play“. For a strategy combination, we name the probabilities for „direct point wins“, „direct point losses“, and „ball in play“ the *success distribution* of the strategy combination.

Let us assume for a moment that the success distributions for all possible strategy combinations are known and do not change during the game. Then, using the theory and algorithms of *Markov chains*, we can calculate for each strategy combination what is the probability of winning for us. Using the theory of *Markov decision problems*, one can even determine a decision rule without simulation, which directly gives one an optimal strategy against any fixed opponent strategy.

However, two hurdles stood in the way of the project at this point: first, it is reasonable to assume that teams will play differently in different match phases play differently, and this is not represented in our basic model. For example, Jonas Reckermann reported that the hitting behavior of teams often changes when the score is close at the end of a set: For example, often the „favorite strokes are applied or also quite deliberately strokes that one did not make earlier in the set and one has saved up, so to speak“. If the style of play depends on the score, then it is not impossible (though not certain) that the success distributions do depend on the score. Second, the success distributions are not known just like that. How could they? If you do something about the first hurdle and make the success distributions dependent on the score, then there are even more probabilities that you don't know, and the second hurdle becomes even higher.

The project team first decided to make progress in overcoming the second hurdle. (Science, after all, proceeds step by step, and you try to understand one thing at a time). So how to get at these success distributions? It quickly became clear that a direct estimate of the relative frequencies of successes and failures in the course of the tournament up to the final was problematic. After all, the probability of winning points „directly depends not only on the skill of the attacking team, but also on that of the defending team. But what if you haven't played the final opponents at all so far in the tournament?

Hoffmeister had the following new idea: The success distribution of an attack is decomposed into the skill distributions of the involved (*skill* is new German for a „skill"). A skill distribution for a strike specifies for all team members the probabilities of executing the strike „exactly as planned", „with deviations", or „totally mishandled". If you execute a hit, the skill distribution depends only on which previous hit you were confronted with and in which quality (this includes both opponent's attacking hits and passes from your own team) and which skills you yourself have for the selected hit. Which stroke is chosen then depends on one's own strategy (e.g. Smash or Shot). In this way, the success distributions that depend on *team pairings* should be computed from distributions that depend only on the *performing individual*.

But how does this work? Hoffmeister's second new idea was to determine the success distributions using a detailed game model of a beach volleyball ball change. Namely, by using „digital twins" (see Chapter 52) with exactly the skill distributions of the real participating individuals in computer simulations to contest a huge number of beach volleyball ball changes. The success distributions are then estimated from the number of „direct point wins", „direct point losses", and the occurrences of „ball in play" in these digital rallies.

The question remains how to get the skill distributions. These are also a lot of probabilities that have to be estimated from previous matches. The advantage is that we can now use every match for the estimation, no matter against whom. For our own team, we could even evaluate corresponding test series in training in large numbers. Since the latter was of course not possible for the project team, video analyses had to be used. This required the lengthy evaluation of video recordings of all matches of Brink/Reckermann and of Alison/Emanuel in the Olympic tournament up to the final.

Olympics 2012

After all, we promised to still resolve the strange strategy designations from the winning schedule in table 1. These are the 16 combinations of the following individual strategies (white = „no", gray = „yes"):

1. Player 1 goes full risk on serve, essentially characterized by *jump serves* (very hard jump serves with a run-up) placed close to the edge of the court, as opposed to *(Jump) Float Serves* (more fluttering serves), which tend to target the center of the court.
2. Player 1 goes full risk in field attack, essentially characterized by *smashes* (very hard smashes, usually placed longline or diagonally close to the sideline or on the „six“ on the back sideline) as opposed to *shots* (placed in an arc across the block into an abandoned corner of the court).
3. Player 2 goes full risk on the serve (analogous to player 1).
4. Player 2 takes full risk on field attack (analogous to player 1).

We can now put ourselves back into the situation before the final, because the winning schedule was created from data that was all available before the final. were available.

If both teams had played the same strategies in the final as they did on average over the course of the tournament up to the final, our analysis suggests that the probability of winning would have been close to 50%. In any case, the exciting course of the actual final does not speak against it. It is still considered an advertisement for the sport of beach volleyball.

But what would our winning schedule have suggested to the Brink/Reckermann team at that time in order to win with the highest possible probability? Here is the pattern again:

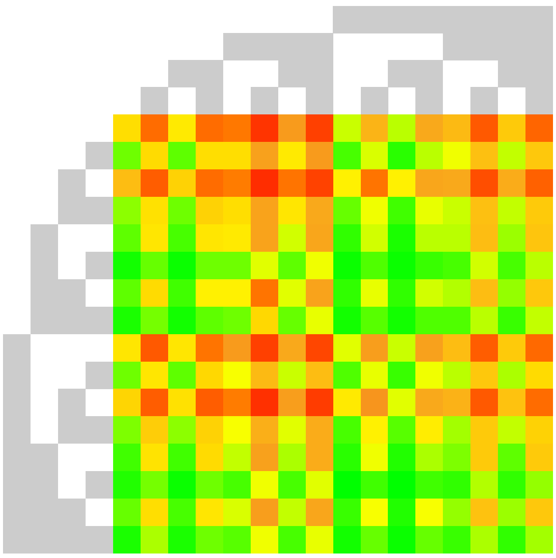


Tabelle 2: Where is the „greenest“ line in the winning schedule?

We are – roughly speaking – looking for a particularly green line. It is interesting how sensitive the human eye is to color differences. We see at a glance that the even rows

2, 4, 6, ... are all significantly greener than the odd rows 1, 3, 5, ... This indicates that player 2 in the German team (Reckermann) should definitely play full risk in the field. And this is regardless of the strategy of the Brazilian team. It is quite similar with the field attacks of Julius Brink, even if he is a top shot player according to Jonas Reckermann's statement. For him, too, the winning schedule would have recommended playing full risk on the attacking shot, since lines 5, 6, 7, 8 and 13, 14, 15, 16 are greener than the other lines.

It looks different for the serves. Whether this stroke is executed with more or less risk makes no visible difference according to the winning schedule for the German team. For player 1 (Brink), for example, you can see that lines 1 to 8 are very similar to lines 9 to 16.

This is not implausible in principle: jump-float serves, which we consider less risky than e.g. super-hard jump-topspin serves, are also sometimes unpleasant to handle when receiving, especially e.g. in windy conditions. You have to score a lot of direct points with the jump topspin or provoke bad attacks to make the risk worthwhile in the final calculation.

We can see that the four lines 6, 8, 14, 16 already look very „green". If we take the principle of mathematical game theory into account, according to which we assume that the opposing team could choose the best strategy against us, then we should still look at the least green field in these lines. And there we notice that in the case of the best answer of the Brazilians (column 6) the 8th row falls out disadvantageously, because the corresponding winning probability of the strategy combination (GER 8, BRA 6) is orange, i.e. less than 50 %. In the other rows, we essentially reach yellow (50 %) against the best answer. This translates to slightly worse if player 1 (Brink) doesn't go full serve, while player 2 (Reckermann) does. We're getting into the realm of estimation error with the color differences here, though.

It is still striking that our numbers predict better chances for Brazil if Alison does not play full risk (i.e. rather float serves) on serve: Columns 14 and 16 (unlike columns 6 and 8) visibly fail to reach 50 % win probability against the best German strategies (the least green field is still green).

If we assume that Germany plays with the strategy whose least green field is still greenest, and Brazil plays with the strategy whose greenest field is least green, then we get a winning probability that arises when both choose their strategies in such a way that they would not deviate from them unilaterally. Such a combination of strategies is called a *pure Nash equilibrium*, after the „inventor" of non-cooperative game theory John Nash (a Hollywood version of his life story is shown in the movie „A Beautiful Mind"). A pure Nash equilibrium need not always exist. If you „mix the strategies (you then play each pure strategy with a certain probability), then it does exist. One then speaks of a *mixed Nash equilibrium*. In our winning schedule, all eligible squares are yellow (50 %). These are (GER 6, BRA 6), (GER 14, BRA 6) and (GER 16, BRA 6), and that promises an exciting final with an open outcome, which is what it turned out to be in the end. Maybe it was a small advantage for the German team to have more different playing styles with similar winning probabilities at their disposal.

You can see very impressively from the winning schedule that at the high level, both teams benefit most from risk in the field attack. Each team would have had an almost certain chance of winning with a risk-taking strategy against a lower-risk strategy of the other team. Against the top teams, restraint is obviously not an option.

But beware: general rules applicable to *all* beach volleyball matches can *not* be deduced from *our* data without additional knowledge. The dependence of the conclusions on the abilities of the teams that meet is, after all, what is new about the approach. It all depends on the details, especially the skill distributions, i.e. how well one team serves or attacks in one way or another, and how well the other takes or blocks the corresponding types of hits. For a different final with different teams, the conclusions might be different. To derive general rules from data, one would use other methods.

More detailed studies on the dependency on skill distributions showed that you have to execute the aggressive strokes with very high consistency to get something out of it. A smash, for example, should be executed with a probability of success of around 90 % to make the risk worthwhile. We were able to observe this from the audience at the 2016 German University Championships in beach volleyball at the University of Bayreuth: In some matches on an intermediate level, the teams that consistently aimed shots into abandoned corners of the court instead of constantly trying to hit smashes from insufficient jumping heights (which is a lot of fun, of course) were able to win.

The limits of our model can be clearly seen if we use the skill evaluation of the final to calculate the win probabilities. This resulted in significantly lower win probabilities for Brink/Reckermann across the board despite the final victory, as if the Brazilians had been able to improve in the final. Jonas Reckermann could name another possible reason for the change: Alison/Emanuel had scored one point more in the overall count in the final because they were able to take the second set very clearly. If the lead in a top-level set is hard to catch, the trailing team may not use all its skills and may even give away a set – and the skills observed in such a phase are no longer representative for other phases of the match. For us, however, these observations count just like all the others. Among other things, there is potential for improvement.

We see that mathematical models are actually never finished. You always find out that there is something you hadn't thought of before. Only by talking to experts can we get to the root of the most important omissions. The exchange between mathematics and application is therefore one of the most important success factors in application-oriented mathematics research.

We would like to thank Julius Brink and Jonas Reckermann very much for their interest in our project and the extremely interesting conversation – and of course still for the terrific Olympic final 2012!

67 Data- and Model-Based Talent Development of Junior Soccer Players**

Prof. Jan Mayer and Dr. Sascha Härtel from TSG Hoffenheim and Oliver Wohak from d-fine GmbH report on a joint project in which mathematical models and machine learning methods were applied to physiological and psychological performance data of junior soccer players of TSG Hoffenheim. The innovative project successfully revealed data-driven strength-weakness profiles and compensation mechanisms, and developed individual development predictions for each player.

Mathematical Models Support Individual Development of Youth Players in Soccer

There is a lot of data in soccer. In the meantime, 3.6 million data points are collected per match on a Bundesliga matchday. However, before a player makes it to the highest German division, he usually passes through the youth academies or junior training centers. And even there – starting with the youngest age groups – the collection of performance data is part of the process. TSG Hoffenheim is a pioneer in the promotion of young talent. In addition to tracking data, which records current performance data for every training session and game, very extensive physiological and psychological performance tests are carried out. The aim is to use this interdisciplinary data to create as holistic a picture as possible of the individual players and teams. The determination of the actual status enables the assessment of the basic requirements (e.g. is a player rather quick-footed or endurance-oriented? is he fast in the head and/or in the legs) as well as a comparison with target values (benchmark values professionals, position-specific requirement profiles).

The strengths/weaknesses profile derived from the tests enables concrete, individual training measures to be taken and, consequently, performance to be optimized. With the help of test repetitions, the long-term development and thus the effectiveness of the chosen interventions can be checked.

Utilization of Performance and Diagnostic Data Supports Talent Development

Compared to other soccer clubs, TSG started structured data collection very early. As a result, extensive, multi-year time series are available for various U-teams on a uniform set of key performance indicators, which can be used for exploratory data analysis as well as for the development of forecasting models. Note that 'U-team' means a team below a certain age. Typical questions in this context include: How do certain metrics change over the course of a career? How do certain correlations develop over time? Which forecasts can be made regarding the transition to the U-teams? All with the aim of identifying patterns in the data that support the individual development of players and enhance their performance potential.

Looking Back - Initial Exploratory Data Analyses Reveal Correlations

Simpler patterns can be identified in the data by means of exploratory analyses. These are, for example, the development of individual attributes over time (univariate analyses) or correlations of different attributes (multivariate analyses). Correlation analyses are particularly helpful when the „similarity“ of historical time series of individual attributes is to be evaluated. Here, the calculation of the correlation gives insight into the strength of the correlation of the attributes as well as the direction: -1 is maximum strongly negatively correlated and +1 maximum strongly positively correlated. Figure 96 shows the correlation values of different attributes from the performance tests.

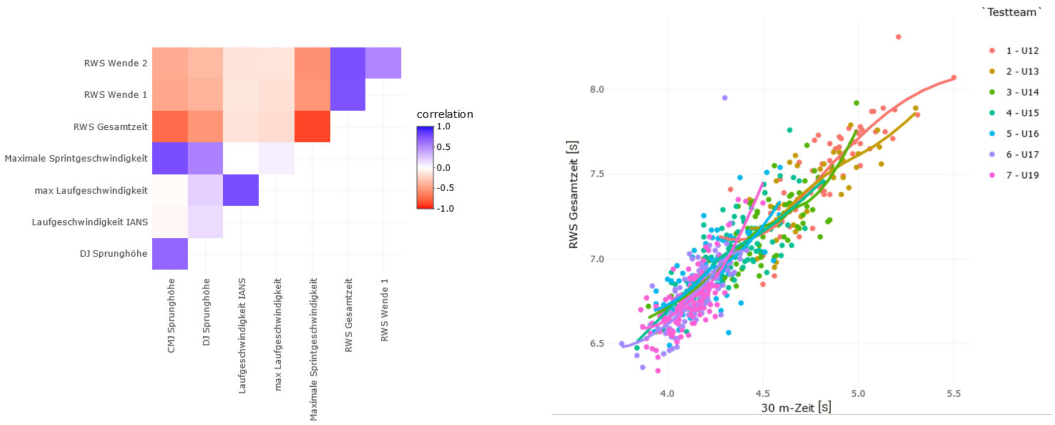


Figure 96: Correlation plot to speed, endurance, and jumping strength scores (left) and bivariate analysis to directional change sprint (RWS) and 30 m sprint time on the right.

For example, it can be seen there that the attributes of jumping power (specifically, here, the various indications of jump height DJ jump height and CMJ jump height) correlate highly with sprint values – strongly negatively with change-of-direction sprint times (RWS) and strongly positively with maximum sprint speed. It can be inferred that players with good speed values typically also have strong jumping power.

In addition to correlation plots, bivariate analyses can help to further investigate correlations between two specific attributes. In the scatter plot in Figure 96, the two attributes RWS total time and 30 m time are compared as an example. Due to the close content of the two attributes, a strong correlation can be seen in the analysis. Players with a fast 30 m time typically also complete the direction change sprint in a short time. However, it can also be seen that this correlation is not perfect. Thus, in an individual case, a player may have a good 30 m time, while performance on the change of direction test is capable of improvement. In summary, the attributes 30 m time and RWS total time seem to be strongly related, but still specific enough to be able to identify strengths and weaknesses in the respective attribute separately. Furthermore, it can still be seen from the U-team specific trend lines that the sprint times stand out from the other teams especially in the U12 to U14 range due to physical development and that the differences between the teams are very small from U15 onwards regardless of age.

Looking into the future – Machine learning models enable predictions about player development

In addition to exploratory data analyses, it is interesting to examine the extent to which prediction models can help make statements about future player development. While explorative analyses merely describe the available data and subsequently show correlations, prediction models enable a prediction of player developments. In our project, this was done by forecasting the transition probability to the next higher U-team, which was determined for each player. This allows early intervention by athletic trainers or sports psychologists to optimally support the players' further development.

For this purpose, established machine learning methods – i.e., mathematics – lend themselves to identifying and exploiting both linear dependencies (e.g., using linear or logistic regression) and more complex relationships (e.g., using tree-based models or neural networks) in the data. These methods have in common that in a process – which is also called „supervised learning") – regularities are detected in historical data, to predict, based on a set of input values (model features), related results or events (model label) as well and reliably as possible based on a set of input values (model features). In this project, a so-called random forest model (a combination of several tree-based models) was trained to predict a U-team-specific transition probability based on players' physiological and psychological test scores. Thus, the features in the model are, for example, physical activity, running speed at the individual anaerobic threshold (basic endurance), or even the month of birth (children born late in the year often drop out of the promotion because they are usually less developed compared to those born early in the year), while the label is the (non)successful transition to the next higher U-team at the end of the season.

The model is first trained with the historical data and then validated. To do this, the existing data is typically separated into a training set (e.g. 80 % of the data) and a validation set (the remaining 20 %). The validation set is used to test the model quality, for example by the number of true-positive and false-positive predictions. This is called the ROC curve (Receiver Operating Characteristics) or a Gini coefficient. or a Gini coefficient (which normalizes the area under the curve to values between 0 and 1). Thus, it can be predicted with a certain reliability that a midfielder of the U17 with 16 years, a physical activity of 1 (maximum), a running speed IANS (12.8 km/h) as well as some other attributes (Figure 97) will make the jump to the U19 with a probability of about 76 %.

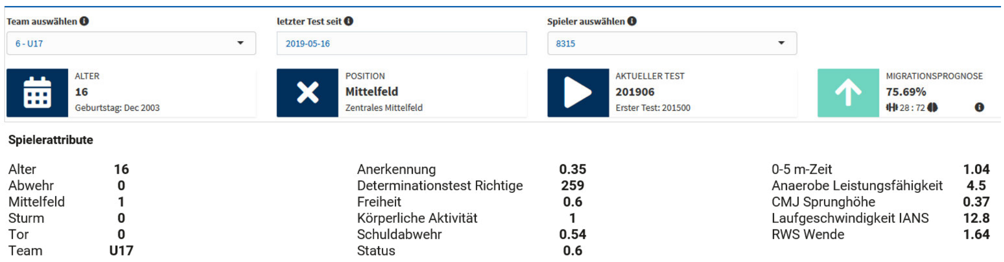


Figure 97: Illustration of the attributes of a U17 player with a high probability of transition.

Making Models Explainable Provides Additional Insights

As a supplement to the coaches' expert opinions, it is helpful to be able to assign a transition probability to individual players and thus to be able to identify data-supported players with high potential at an early stage. In addition, however, in order to be able to provide each player with the best possible individual support, we also need to know which characteristics are strongly pronounced in players with high potential and which physical or mental characteristics in particular should still be trained in „weaker " players. By default, machine learning models cannot make a direct statement about this. However, there are mathematical methods that allow quantifying the contribution of each attribute to the prediction outcome. This is then called „explainable AI" (xAI) or explainable artificial intelligence. One of these methods – which was also used in this project – is SHAP (Shapley Additive exPlanations). It builds on an idea originally from game theory and seeks to quantify the marginal contribution of each player in a cooperative game with a measurable outcome. Applied to machine learning methods, SHAP quantifies the marginal contribution of each model feature on the model prediction. This is achieved by comparing the outputs of many similar models. These models do not differ in their parameterization, but cover all possible combinations of feature compositions. By comparing the respective model forecasts, it is possible to back-calculate what contribution each feature has on the forecast of the full model. In particular, it is thus possible in our model to show which features tend to drive or depress the transition probability for a given player – compared to the average player. This is illustrated by the „force plot in Figure 98, which shows that the player shown stands out in particular for his physical activity and good performance in the determination test (cognition test). One characteristic that the player can still work on to improve further, and where he performs worse than average, is his baseline endurance – shown here as running speed at the individual anaerobic threshold (IANS).

From the xAI approach and the representation by means of „force plots", data- and model-based strength-weakness profiles for the players can thus be derived – also shown again in Figure 99. In addition, other interesting findings can be derived that can be exploited by performance diagnosticians and sports psychologists. For example,

Figure 98: Force plot that shows which attributes distinguish the represented player compared to an average player (base value).

compensation effects were analyzed in the project. These indicate whether a player's strengths are more physiological or psychological. In the case of the U17 player shown above, the ratio of physiological to psychological strengths is 28:72. This means that the player has pronounced mental abilities and compensates for his athletic deficits through them. In particular, it can be seen across the different youth teams that physiological characteristics are more important for promotion at a younger age, but that psychological skills become all the more important at an older age for the leap into the professional squad.

Figure 99: The strength-weakness profile and the compensation effect (dumbbell = physiological, brain = psychological).

Data and Models are Becoming Increasingly Important in Soccer as Well

Overall, the project has shown that mathematical methods and models can help to exploit existing data in soccer. Using historical data on youth players of TSG Hoffenheim, a machine learning model was implemented that derives transition probabilities for individual players and also enables the definition of strength-weakness profiles and compensation effects. Thinking further, this type of data can also be used in soccer to identify, for example, injury susceptibility or overuse – for example also in combination with tracking data from training and matches.

68 The Optimal Free Throw in Basketball**

When a basketball player, such as Shaquille O'Neal or Dirk Nowitzki, goes up for a free throw, he probably doesn't think about how to choose the angle and speed of the release of the ball; instead, he intuitively throws the ball toward the basket to maximize the probability of a successful basket.

In this chapter we will use the free throw in basketball to explain the process of mathematical modeling and thus answer the question „How can we improve the scoring rate of the free throw?“ After a modeling of the considered mostly complex process under consideration, it is often necessary to simplify the model (e.g. to get solutions in real time). Subsequently, one has to discuss methods for the numerical computation of approximate solutions (here: derivative-free multi-objective optimization algorithms). Furthermore, at the end we want to explain how to optimize zone defense in basketball with the help of so-called *Voronoi diagrams*.

Other important questions would be: how does air resistance affect the trajectory of the ball? Where is it actually best to aim; at the front or back of the ring or in the middle of the basket or ...? Which error (i.e. deviation from the optimal parameters) can one allow oneself in order to still achieve a basket success? So, in the following we consider a basketball player of a certain size and we want to answer these questions.

During a free throw, the basketball leaves the player's throwing hand with a certain release angle and a release velocity (and possibly a spin); the further trajectory is described by equations of motion. The success of the basket can happen in different ways: with/without touching the ring, after bouncing off the so-called backboard or after jumping back and forth between parts of the ring and the backboard several times.

We make some assumptions to arrive at a simple model. We only consider throws where the basketball hits the net directly or hits the backboard of the ring first and then goes directly into the net. We neglect the influence of air resistance and also spin (because we don't consider bounces off the backboard. Furthermore, we neglect lateral errors in the trajectory, so that we have equations of motion in two dimensions (height, distance) are sufficient.

With these assumptions, one can describe the trajectory of the basketball at a given throw-off angle θ^0 by means of differential equations and, after some computation[68], obtain a statement of which launch angle θ^0 allows the largest possible deviation. For this purpose, we denote by θ^0_{min} and θ^0_{max} the smallest and largest deviating angle θ, respectively, at which the free throw will just result in a hit, i.e. in case of a successful throw the angle θ^0 must be in the interval $[\theta^0_{\text{min}}, \theta^0_{\text{max}}]$.

Mathematically, a *objective function*

$$e(\theta^0) = \min\{(\theta^0 - \theta^0_{\text{min}}), (\theta^0_{\text{max}} - \theta^0)\}$$

[68] J.M. Gablonsky, A.S. Lang, *Modelling Basketball Free Throws*, SIAM Review 47 (2005), 775-798.

and define the problem as a so-called *max-min-problem*: Determine that optimal drop angle θ^0_{opt} for which the objective function $e(\theta^0)$ is maximized. This objective function depends, of course, on many other parameters, e.g. dropping speed v^0, size of the player h_s, and so on. By the way, this objective function $e(\theta^0)$ is continuous but not everywhere differentiable, because it obviously has a „kink". It turns out that it is better to optimize ejection velocity and angle simultaneously, this is called *multi-objective optimization*.

Figure 100: A lecture about the optimal free throw at the basketball second league club RSV Stahnsdorf

In zone defense in basketball (or space coverage in soccer), one would rather like to solve a min-max problem: the maximum distance between defenders and opposing offensive players should be minimized. For this purpose there is a simple mathematical tool, the *Voronoi diagram*. Here the playing field is divided into areas (so-called cells), which are determined by given points (so-called centers, i.e. the players). Each such Voronoi cell has exactly one center and includes all points of the playing field which (with respect to the Euclidean norm) are closer to the center of the cell than to any other center. From all points that have more than one nearest center and thus form the boundaries of the cells, the Voronoi diagram is formed.

One can imagine a player completely 'owning' the point of the basketball court on which he is standing, having no control over a point very far away, and having little control over a point nearby (e.g. 1.5 m away) – except perhaps when another player is standing even closer to that point. On the playing field, regions have different values (as in real estate) and players act rationally, i.e. they swap their playing field regions as they move (and move the ball, opponent) in such a way that it is a winning strategy for their team.

In this sense, Cervone, Bornn, and Goldsberry introduced a „weighted Voronoi" approach to define each player's court ownership at any given point in time[69]. They first

[69]D. Cervone, L. Bornn, K. Goldsberry, *NBA Court Realty*, MIT Sloan Sports Analytics Conference 2016,

divided the half court into $M = 576$ cells of equal size, each about $\times$ 60cm. For player i, $X^i(t)$ denotes an M-dimensional vector representing its possession value in each of the M playing field cells at time t. The m-th entry of $X^i(t)$ is inversely proportional to the distance ('dist') between player i and playfield cell m at time t only if no other player is closer to playfield cell m:

$$w^i_m(t) = \text{dist}(\text{player}\, i, \text{playfield cell}\, m) \quad \text{at time } t$$

$$X^i_m(t) = \begin{cases} \frac{1}{1+w^i_m(t)}, & i = \text{argmin}_j\, w^j_m(t) \\ 0, & \text{else} \end{cases}$$

Thus, space is divided among players according to the Voronoi diagram of player positions, and within each segment they control, players' investment in space is inversely proportional to their distance from that space. In particular, an offensive player's possession of space implicitly encodes information about the defensive player's positioning. As a defender approaches player i, the Voronoi partition of player i decreases, implying that more entries in the pitch possession vector $X^i(t)$ are zero.

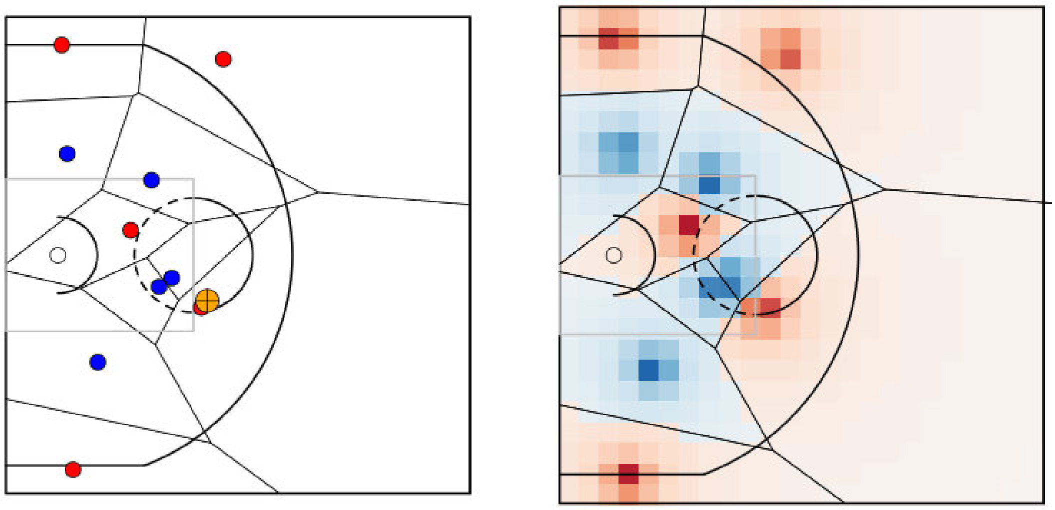


Figure 101: Example of a map for 'ownership' of playfield cells. A player's control over each playfield cell in his Voronoi segment (left) is inversely proportional to his distance from that cell (right).

By the way, these Voronoi diagrams are also used in computer games to analyze the game situations.

http://www.lukebornn.com/papers/cervone_ssac_2016.pdf.

69 The Secret Behind a Successful Finish in Cycling*

What is the secret behind a successful final sprint in cycling? Not just a top sprinter, but also a calculation that determines the optimal sprint move.

Two of the best sprinters in the current cycling peloton ride for the Dutch team Giant-Shimano: the Germans Marcel Kittel and John Degenkolb. Kittel excels in flat arrivals in flat stages. Degenkolb has sprints to contend with in heavier races with more climbs. In both the 2013 and 2014 Tour de France, Kittel won no fewer than four stages, including the prestigious final stage in Paris twice. Degenkolb's wins included the classic Paris-Tours in 2013 and a stage of the Tour of Italy, and the classic Ghent-Wevelgem in 2014. Movement scientist Teun van Erp is employed by Giant-Shimano and analyzes, among other things, the sprints of Kittel and Degenkolb.

Van Erp receives all the measurement data from his riders, both from training sessions and races: „This so-called SRM data tells exactly what power a rider delivers at what moment. From this I can see how long a rider can maintain a certain power and how good his form is. I analyze the data with special software and apply all kinds of mathematical tricks."

He also uses this data to determine the optimal sprint move. Van Erp: "Theoretically, the optimal sprint train is determined by how long a rider can maintain a certain power level in the final. In a typical flat finish, the first man on our sprint train is in the lead with one or three kilometers to go. He slowly increases the speed of the peloton over a kilometer and a half so that no one can escape. He is selected for one to one and a half minutes against the fifty kilometers per hour. After that, the numbers two and three of the sprint train are in front, both about five hundred meters behind each other. They increase the speed even more, up to fifty-five kilometers per hour. With about five hundred meters to go, the 'lead-out' is in the lead, with the sprinter in his wheel. The lead-out can also sprint well and gets our lead sprinter going. That sprints towards the finish at a speed of more than seventy per hour and with a peak power of about two thousand watts."

Figure 102: Graph of power (lower line) and speed (upper line) as a function of distance to the finish (for five riders from the sprint team.

70 Cyberpunk 2077: The Math Behind Computer Graphics**

Linear algebra provides the foundation for lighting models and three-dimensional motion in computer graphics.

In this chapter we want to investigate to what extent mathematics plays a role in computer games. We will get to know how vector calculus (linear algebra) helps us to characterize objects (such as walls in a maze) in three-dimensional space and helps these objects to move, to rotate. Mathematics is also needed to calculate which parts of an object are visible in the first place and how lighting affects the representation of the surface.

The basic idea of 3D graphics is to take a mathematical description of a world and turn it into an image of what that world would look like to someone inside it. The mathematical description could be like this: there is an object box with center (1,2,3), side length 2 and color yellow. The observer (exactly his eyes) is at (5,14,9) and looks directly at the center of the box. With this you can calculate how the world would look like for this observer.

Figure 103: Cover image Cyberpunk 2077 (Source: CD Projekt Red, Press Center)

Now take an observer with eyes at point A; between him and the box is a glass plate (plane) on which he wants to paint the box he sees. One of the box corners is at point B, and to find out where this corner should be on the glass plate, draw a line L from his eyes A to the box corner B, and then see where this line passes through the glass plate G,. Such an intersection of straight line with plane can be easily determined. If you follow this procedure for every place of the box, you get an image of the box on the glass plate, i.e. the screen. And this is exactly what happens very roughly in the computer, for example, when you walk around in a game like Cyberpunk 2077, although the details are a bit more complicated.

The Ray Tracing Process

What was just described above is similar to what the computer does every time you walk around in the game like Cyberpunk 2077, although the details are slightly different. The latest computer games use more complicated descriptions of the world, using curved surfaces, NURBS, and other strange sounding things, but in the end it always reduces to triangles.

Let's assume that we have triangulated our object into triangles. The next step is the illumination model: the amount of light or shadow that falls on each triangular surface depends on the orientation of that triangle with respect to the light source. If a triangle faces the light source directly, it will be brightly illuminated, but if it faces away from the light source, it will be in heavy shadow. The orientation of each triangular surface is described by a normal vector pointing perpendicularly out of the respective planar surface. Similarly, the rays of the light source can be described by a vector, and the scalar product of these two vectors gives the cosine of the angle between them.

At this point we want to briefly introduce the *scalar product*. To calculate the scalar product of two vectors $u = (x_1, y_1, z_1)$ and $v = (x_2, y_2, z_2)$, multiply their respective coordinates and add them:

$$\langle u, v\rangle = x_1x_2 + y_1y_2 + z_1z_2.$$

The scalar product can also be written as the product of the length of the two vectors and the angle θ between them:

$$\langle u, v\rangle = (\text{length of } u) \cdot (\text{length of } v) \cdot \cos\theta.$$

If a triangle faces the light source directly, the angle between the normal and the direction of the light rays is $0°$, and the scalar product of these two vectors gives the value 1, i.e. the maximum amount of light falls on this surface and it should therefore be shaded white. If a triangular surface is almost lateral to the light rays, so that the angle between the surface normal and the light direction is close to $90°$, the value of the scalar product is close to 0, so the triangle should be shaded very dark because it is in shadow. If a triangle is away from the light, so the angle is greater than $90°$, then the scalar product becomes negative and this value is normally set to zero by the computer. Now add to this the fact that all 3D objects are moving. Thus, the computer must constantly repeat these (relatively simple) calculations for all triangles (the so-called *rendering*). For this purpose, there is now specialized hardware that does these calculations. Often the triangulation is also coarsened to make the calculations more effective for fast moving objects; the human eye cannot resolve so finely anyway.

The next step is to make the surfaces appear more realistic without further increasing the number of triangles. This is done with the concept of texture. For applications that

Figure 104: Screenshot 2077 (Source: CD Projekt Red, Press Center)

require faster rendering time, such as real-time rendering in computer games, one uses *normal mapping*. It corresponds to the method described before, where the texture with its unevenness is obtained by locally high-resolution models.

For applications, such as movies, that allow long times for rendering individual frames, there is another method. One tries to use *displacement mapping* with positive and negative values to encode exactly how far the points on the surface of the simplified grid differ from the points on the original grid.

71 Sudoku – Fun, but also Serious Math*

Math helps solve even the most difficult Sudoku puzzles.

Solving Sudoku has been a very popular pastime since 2004 and can rightly be called the puzzle of the 21st century. People with Sudoku puzzle books can be seen in waiting rooms, on trains and on airplanes everywhere. What you need to do to solve a Sudoku puzzle can be explained in a minute. But how to do it is another story. There are easy Sudokus, and there are very difficult ones. The puzzle books you can buy state how difficult the puzzles are, and the most difficult puzzles are rated 10, 11 or even 15 stars.

Mathematicians are also very interested in Sudoku. They try to create very difficult Sudokus that are almost impossible for a human to solve. The following puzzle is currently the most difficult in the world:

				9			5	
	1						3	
		2	3			7		
		4	5				7	
8						2		
					6	4		
	9			1				
	8			6				
		5	4					7

This puzzle can be solved using the techniques described in the book „Solve Any Sudoku“, but it requires a lot of experience and will still take you a whole day of work.

For mathematicians, however, there are many other interesting questions. For example: What is the minimum number of digits that should be prescribed to ensure that a Sudoku is uniquely solvable? The answer is 17 and was proven in 2012 by an Irish mathematician, Gary McGuire, but it took him 7 million hours on a powerful computer system!

Another interesting question is: How many different Sudokus are there? „Distinguishable“ means that at least one number in the puzzle is different; symmetry relationships (such as rotations) are not considered; they count as different. In a 2005 study, Bertram Felgenhauer and Frazer Jarvis calculated the number of distinguishable Sudoku puzzles by mathematical means (group theory) and arrived at the number

6,670,903,752,021,072,936,960, confirming the value that Günther Stertenbrink had already postulated in 2003. Bertram Felgenhauer is a very talented mathematician who won the silver and gold medals of the International Mathematical Olympiad in 1995 and 1996, respectively.

Mathematicians often use Sudokus to test their mathematical algorithms, for example in graph coloring problems. Interested in more mathematical facts about Sudoku? See the website https://www.loesejedessudoku.de.

72 Deal or no Deal? The Secret Strategy of the Bank**

How can math be used to analyze a „secret“ strategy?

In the TV show „Deal Or No Deal“ or „Millionendeal“, the contestant chooses one of 26 (or 20) suitcases, each containing a sum of money between 0.01 € and 2,000,000 €. All the money amounts are displayed on a scoreboard in the TV studio, but you don't know which amount is in which suitcase. The game begins with the candidate choosing a suitcase, which initially remains locked. Then, one after the other, the candidate selects other suitcases, which are opened and whose amount disappears from the scoreboard, since their amount cannot be in the locked suitcase.

In the course of the game, after a certain number of suitcases have been opened, the „bank“ makes the candidate an offer (a „deal“) if he sells his selected suitcase to them. This offer is based on the amounts of money still on the scoreboard. The more higher amounts still in play, the higher the bank's offer. It seems as if the bank's offer is (somehow) based on the average of the amounts of money still in play (the expected value of the game at that point).

If the candidate accepts this offer, the game is over and he receives this amount of money from the bank. If he rejects the deal, the game continues and more suitcases are opened and more offers are made by the bank. If the candidate rejects all the bank's offers, he receives the amount of money from the originally selected suitcase. There are also variants where (if the candidate makes his deal before round 7) the game is played to the end in the hypothetical, i.e. the remaining rounds with the offers are played as if the game was not over yet.

The question here is whether it is a fair competition or whether the show is „cheating its participants. One can use simple mathematical terms, such as expected value and equalization, to examine whether there are irregularities or special patterns.

Bowling Green State University mathematics student Daniel R. Shifflet, in his article[70] explored this very question and found the following. In the rounds of the game that were still in progress, initial calculations showed that the banker offered participants amounts of money far below their expected value. On the other hand, in the hypothetical rounds described above, these offers by the banker were close to or even above this expected value of the game. So, had the banker played fair or „played around with the numbers“?

Daniel R. Shifflet divided the data into the two situations to be tested:

1. official offers to participants who are still in the game.
2. hypothetical offers to participants who have accepted a deal.

Since participants with large amounts of money in play would obviously get larger offers than participants with only small amounts of money left (official or hypothetical),

[70] D.R. Shifflet, *Is Deal or No Deal Cheating Its Contestants?*, Ohio Journal of School Mathematics (63) (2011), 5-10. https://kb.osu.edu/bitstream/handle/1811/78122/OJSM_63_Spring2011_05.pdf

it is better to express these offers as a percentage of the expected value of the game outcome. This would give us a value that is independent of whether the contestant had a „lucky day“ or not. If the bank plays „fairly“, the averages of these percentages should be about the same round by round for the official and hypothetical situations.

In an example data set, Daniel R. Shifflet found that participants who are still in the game receive offers from the bank that average about 76 % of their expected value, while participants who are only hypothetically playing receive offers of about 96 % of the same expected value. This 20 percentage point difference seems to indicate that the bank does not follow the same rules for making offers when the game is still active and when it is hypothetical.

Given the large number of data sets, Daniel R. Shifflet suspected a linear relationship between the expected value of the game and the bank's bids. Using a balancing calculation, he found a straight line that best fit this data. This produced a model for how the bank arrives at its bids. His model produced the following results. For participants still in play in round 6, the bank would make offers equal to 65.11 % of the expected value. On the other hand, for participants only hypothetically playing the game in round 6, his model (after removing a so-called outlier value) yields an offer of 94 % of the expected value, i.e., a drastic increase over the earlier situation.

Given all this information, there is strong evidence that the banker deviates from his formula for calculating deals (whatever that is) during hypothetical rounds of the game. He appears to be offering more money than he would have if the participants were still actually playing and able to take the money.

Acknowledgements

We would like to take this opportunity to thank everyone who contributed to this book

- José Acosta Fernandez, Wuppertal Institut für Klima, Umwelt, Energie
- Dr. Maria Barbarossa, Frankfurt Institute for Advanced Studies (FIAS)
- Dr. Petra Bauer, Siemens AG, Technology, München
- Robin Befeldt, Polizei Berlin
- Dr. Andreas Binder, MathConsult GmbH, Linz
- Prof. Dr. Ralf Borndörfer, FU Berlin und Zuse-Institute Berlin
- Prof. Dr. med. Helmut Brunner, Universität Düsseldorf, Bergische Universität Wuppertal
- Dr. Yao-Li Chuang, California State University, Los Angeles
- Dr. Ágnes Cseh, Hasso-Plattner-Institut, Potsdam
- Sören Dittmer, Universität Bremen
- Prof. Dr. Maria Rita D'Orsogna, California State University, Los Angeles
- Dr. Martin Friesen, Dublin City University
- Daniela Gaul, Bergische Universität Wuppertal
- Dr. Sascha Härtel, TSG Hoffenheim
- Dr. Thomas Hahn, Chief Expert Software, Siemens AG, Nürnberg
- Dr. Dirk Hartmann, Senior Principal Scientist, Simulation and Digital Twin, Siemens AG, München
- Dr. Markus Hinterkausen, Zentralbereich Forschung und Vorausentwicklung, Robert Bosch GmbH
- Leon Hoffe, Bergische Universität Wuppertal
- Dr. Susanne Hoffmeister, InVision AG, Düsseldorf
- Dr. Markus Hoschek, HEAG Holding AG, Darmstadt
- Prof. Dr. Wilhelm Huisinga, Universität Potsdam
- Dr. Uwe Iben, Chief Expert Applied Mathematics, Robert Bosch GmbH, Renningen
- Dr. Markus Kantner, Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin
- PD Dr. Jörg Kienitz, Bergische Universität Wuppertal, Quaternion, AcadiaSoft
- Prof. Dr. Kathrin Klamroth, Bergische Universität Wuppertal
- Dr. Margrit Klitz, Deutsches Zentrum für Luft- und Raumfahrt (DLR), Köln
- Tobias Kluth, Universität Bremen
- Dr. Thomas Koprucki, Weierstraß-Institut für Angewandte Analysis und Stochastik, Berlin
- Dr. Igor Kossaczký, Aptiv, Wuppertal
- Prof. Dr. Ekaterina Kostina, Ruprecht-Karls-Universität Heidelberg

- Christine Krüger, Wuppertal Institut für Klima, Umwelt, Energie
- Mira Küpper, Bergische Universität Wuppertal
- Prof. Dr. Stefan Kurz, TU Darmstadt und Robert Bosch GmbH
- Benjamin Leonhardt, Bergische Universität Wuppertal
- Johannes Leuschner, Universität Bremen
- Prof. Dr. Peter Maass, Universität Bremen
- Corinna Maier, AbbVie Deutschland GmbH & Co. KG
- Prof. Jan Mayer, TSG Hoffenheim
- Prof. Dr. Nicole Megow, Universität Bremen
- Prof. Dr. Volker Mehrmann, TU Berlin, Präsident European Mathematical Society
- Dr. Michael Meister, Parlamentarischer Staatssekretär, Bundesministerium für Bildung und Forschung
- Dr. Georg Mogk, Principal Expert Applied Mathematics, Bayer AG, Leverkusen
- Dr. Thomas Mrziglod, Head of Applied Mathematics, Bayer AG, Leverkusen
- Niels Neveling, Senior Consultant, B&W Deloitte GmbH
- Prof. Dr. Jörg Rambau, Universität Bayreuth
- Ronan Richter, Universität Bayreuth
- Dr. Kim Rossmo, School of Criminal Justice & Criminology, Texas State University
- Dr. Matthias Rottmann, Bergische Universität Wuppertal
- Prof. Dr. Helga Rübsamen-Schaeff, ehem. CEO von AiCuris GmbH, Wuppertal, Nationale Akademie der Wissenschaften Leopoldina
- Dr. Guillaume Sagnol, TU Berlin
- Dr. Hanna Schilar, Bergische Universität Wuppertal
- Maximilian Schmidt, Universität Bremen
- Prof. Dr. Anita Schöbel, Institutsleiterin des Fraunhofer-Institut für Techno- und Wirtschaftsmathematik (ITWM), Kaiserslautern
- Prof. Dr. Ingo Scholtes, Bergische Universität Wuppertal
- Dr. Britta Schulze, Bergische Universität Wuppertal
- Prof. Dr. Andreas Schuppert, RWTH Aachen und Bayer AG
- Prof. Dr. Armin Seyfried, Bergische Universität Wuppertal
- PD Dr. Michael Stiglmayr, Bergische Universität Wuppertal
- Matthias Stursberg, R&D Manager Vaillant GmbH, Remscheid
- Tobias Suszka, Bergische Universität Wuppertal
- Alexander Tesch, Zuse-Institute Berlin
- Jun.-Prof. Dr. Antoine Tordeux, Bergische Universität Wuppertal
- Dr. Uwe Vollmer, Zentralbereich Forschung und Vorausentwicklung, Robert Bosch GmbH

- Dr. Utz Wever, Senior Principal Scientist, Simulation and Digital Twin, Siemens AG, München
- Oliver Wohak, d-fine GmbH, Frankfurt am Main