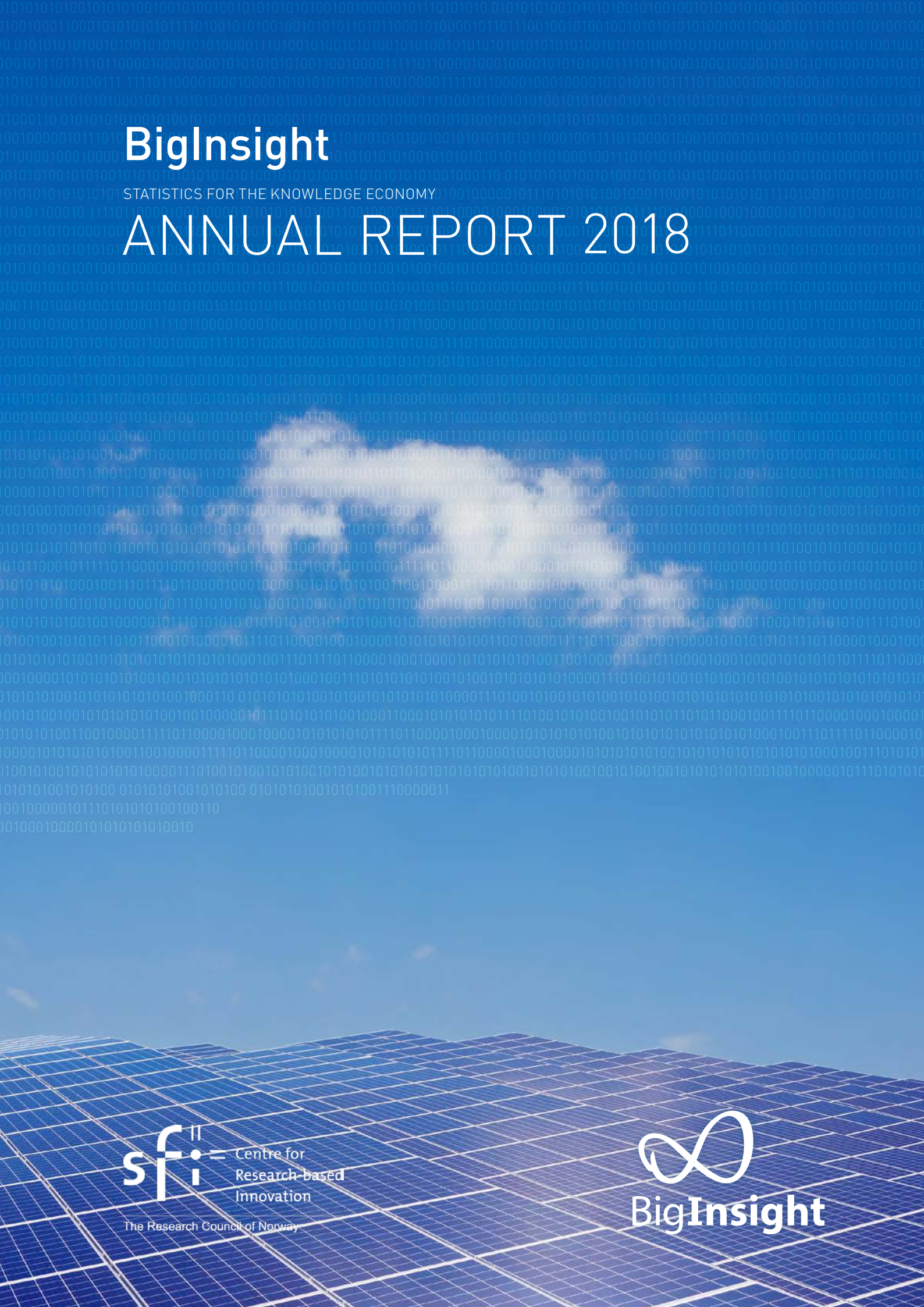
BigInsight

STATISTICS FOR THE KNOWLEDGE ECONOMY

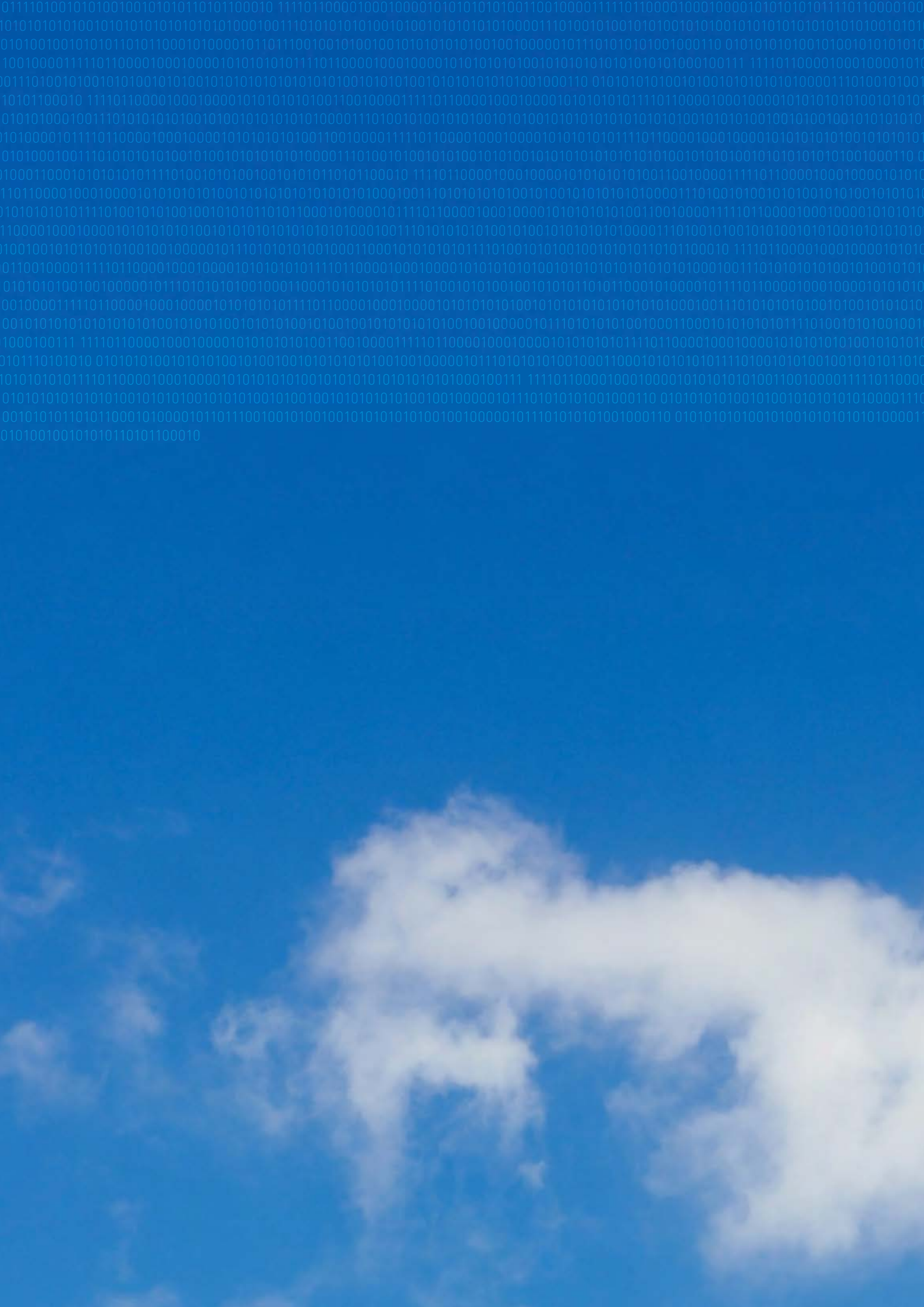
ANNUAL REPORT 2018

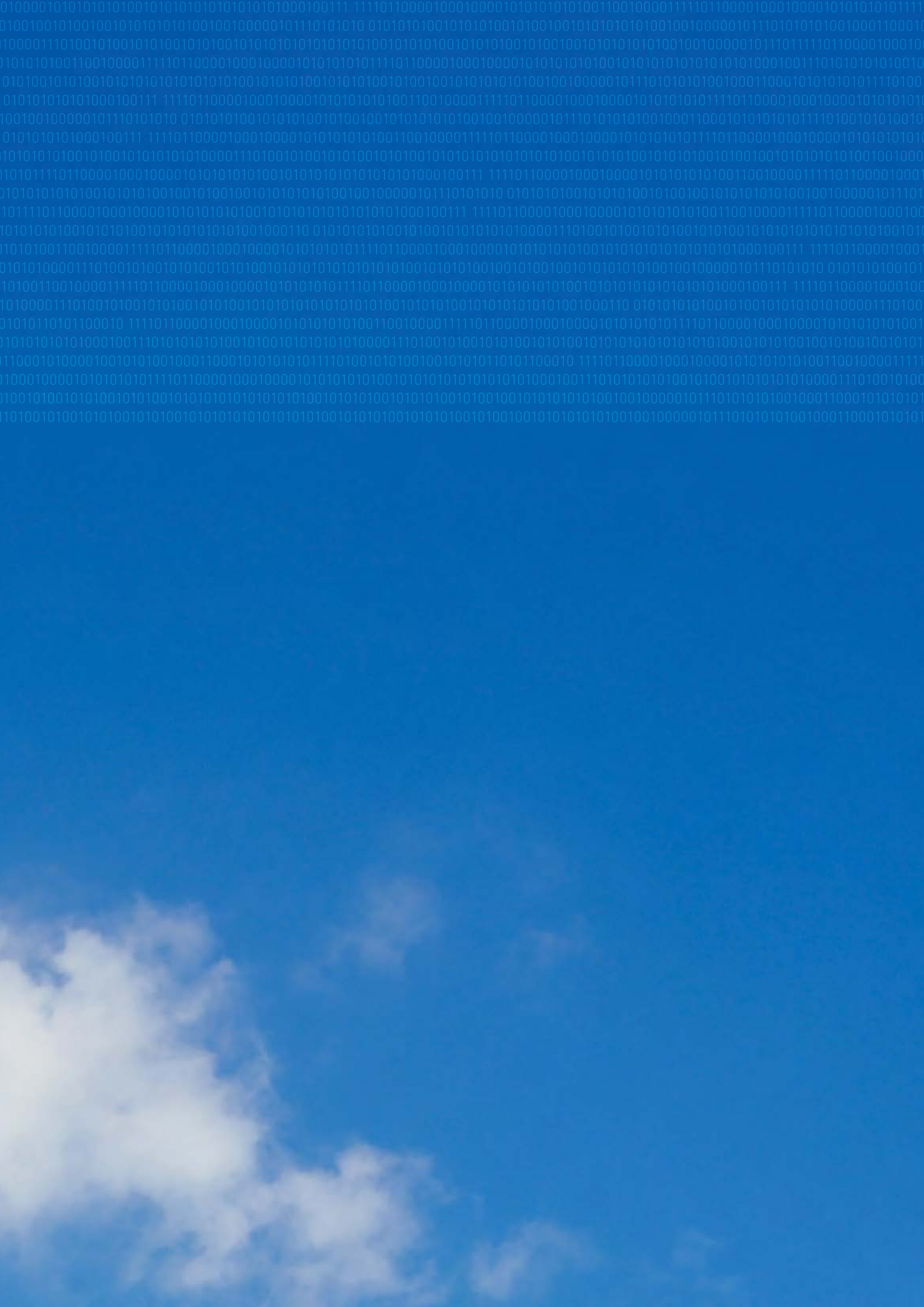


Centre for Research-based Innovation

The Research Council of Norway

BigInsight





"Between data and insight there are methods. BigInsight develops statistical and machine learning methods which are instrumental for progress."

Ingrid Glad, professor and co-director of BigInsight

CONTENT

SUMMARY

BigInsight is a Norwegian centre for research-based innovation, funded by the Norwegian Research Council and a consortium of private and public partners. We produce innovative solutions for key challenges facing our partners, by developing original statistical and machine learning methodologies. Exploiting complex, huge and unique data resources and substantial scientific, industrial and business knowledge, we construct personalised solutions, predict dynamic behaviours and control processes that are at the core of the partners' innovation strategies. Digitalisation of the Norwegian industry and society benefits from BigInsight that produces powerful instruments for the analysis of data.

We discover radically new ways to target products, services, prices, therapies and technologies, towards individual needs and conditions. This provides improved quality, precision, value and efficacy. We develop new approaches to predict critical quantities which are unstable and in transition, such as customer behaviour, patient health, electricity prices, machinery condition. This is possible thanks to the unprecedented availability of large scale measurements and individual information together with new statistical theory, computational methods and algorithms able to extract knowledge from complex and high dimensional data.

When we develop methods and algorithms we consider five principles: responsibility, explainability, accuracy, auditability and fairness.

Research at BigInsight will lead to value creation and will support our partners' leading position.

In the era of digitalization, BigInsight produces competence and capacity for the Norwegian knowledge-based economy, contributing to the development of a sustainable and better society.

This is the annual report of the fourth year of BigInsight. Innovation results are highlighted, together with the broad spectrum of research projects.

"AI and digitalization are recognised as key factors for development and progress of our world. Statistical analysis and uncertainty quantification of predictions and estimates are crucial for good decision making. BigInsight is able to produce the methods and the algorithms that are needed."

Arnoldo Frigessi, director of BigInsight

VISION AND OBJECTIVES

Fulfilling the promise of the big data revolution, the center produces analytical tools to extract knowledge from complex data and delivers biginsight. Despite extraordinary advances in the collection and processing of information, much of the potential residing in contemporary data sources remains unexploited.

Digitalisation means producing data, organizing and storing data, accessing data and analyzing data. BigInsight works in the latter axis of digitalisation. There is a dramatic scope for industries, companies and nations – including Norway – to create value from employing novel ways of analysing complex data. The complexity, diversity and dimensionality of the data, and our partner's innovation objectives, pose fundamentally new challenges to statistical inference. We develop original, cutting-edge statistical, mathematical and machine learning methods, produce high-quality algorithms implementing these approaches and thereby deliver new, powerful, and operational solutions. Our solutions aim to be explainable, fair and responsible.

BigInsight's research converges on two central innovation themes:

- personalised solutions: to move away from operations based on average and group behaviour towards individualised actions
- predicting transient phenomena: to forecast the evolution of unstable phenomena for system or populations, which are not in equilibrium, and to design intervention strategies for their control

Our solutions are significantly better than the state-of-the-art, thanks to courageous and creative methodologies that extract knowledge from structure in complex data and integrate data from various sources.

Our research is open: we publish generic methodology and their new applications in international scientific journals.

Through training, capacity building and outreach, BigInsight contributes to growth and progress in the private and public sector, in science and society at large, preparing a new generation of statisticians and machine learners ready for the knowledge based economy of the future.

"We believe data science, statistics and machine learning will change the world. We are still pioneers; we educate the next generation of professionals and leaders; we join the public conversation; we move the boundaries of knowledge for the public good."

Arnoldo Frigessi, director of BigInsight

Personalised solutions

The core operation of our partners involves interacting with many individual units: at Telenor, for example, millions of individual mobile phone customers are part of a communication network; at Gjensidige, a million policyholders share risks of contingent, uncertain losses; for DNB, customers transfer money and receive loans; at OUS, cancer patients need to be treated in the most effective personalized way; for DNV GL and ABB, hundreds of sensors register the functional state and operation of a vessel at sea. There are many common characteristics:

- a high number of units/individuals/sensors;
- in some cases, massive data for each unit; in other cases, more limited information;
- complex dependence structure between units;
- new data types, new technologies, new regulations are available;
- in most cases, units have their own strategies and are exposed to their environment.

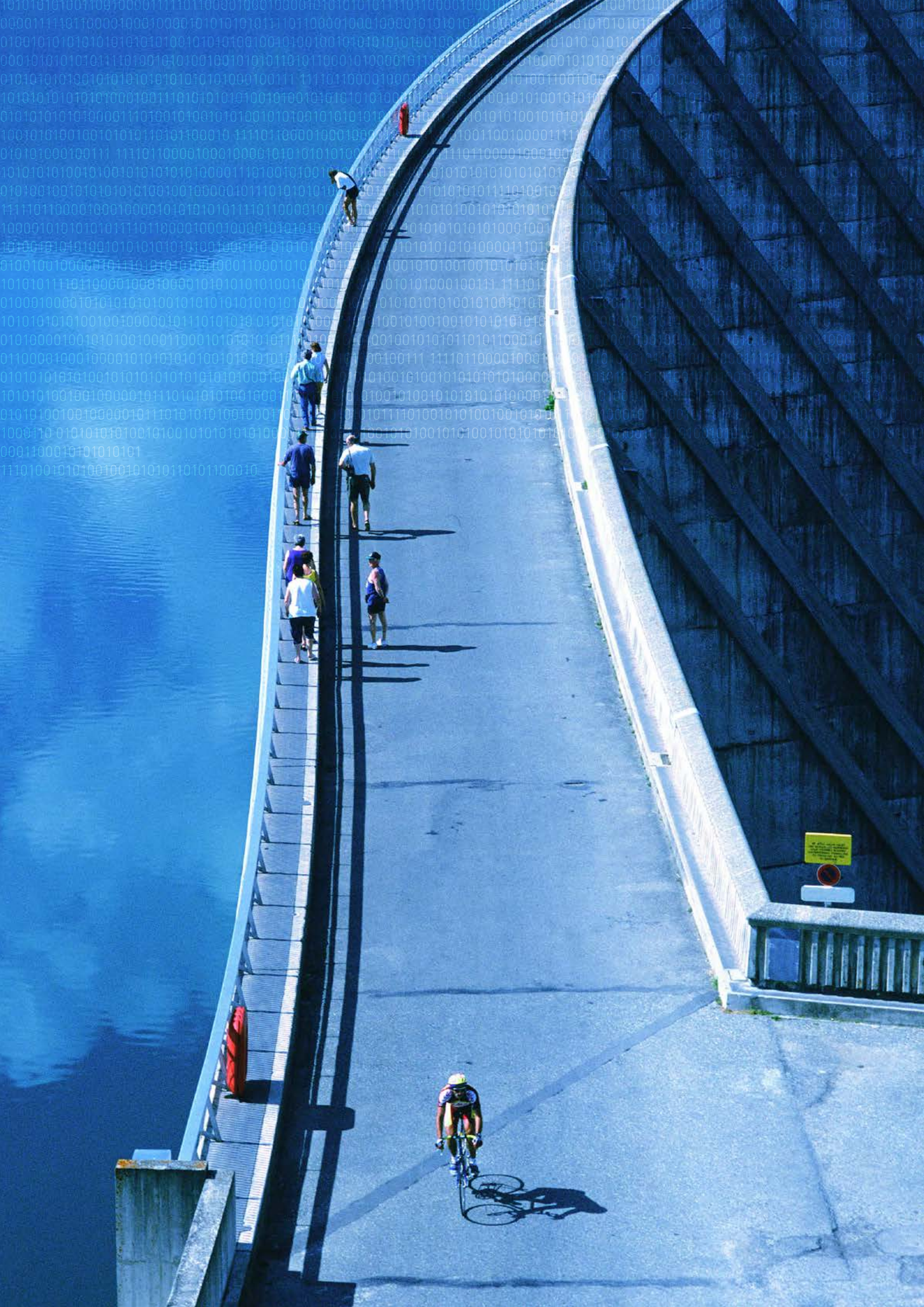
Each partner has specific objectives for and with their units, but they share the goal to fundamentally innovate the management of their units, by recognising similarities and exploiting diversity between units. This will allow personalised marketing, personalised products, personalised prices, personalized recommendations, personalised risk assessments, personalised fraud assessment, personalised screening, personalised therapy, sensor based condition monitoring, individualised maintenance schemes, individualised power production and more – each providing value to our partner, to the individuals and to society: better health, reduced churn, strengthened competitive-ness, reduced tax evasion, improved fraud detection and optimised maintenance plans.

Predicting transient phenomena

The modern measurement instruments, the new demands of markets and society and a widespread focus on data acquisition, is often producing high frequency time series data. As never before, we are able to measure processes evolving while they are not in a stable situation, not in equilibrium. A patient receiving treatment, a sensor on a ship on sea, a customer offered products from several providers, a worker who lost his job, the price of an asset in a complex market – all examples of systems in a transient phase. DNB, NAV, Skatteetaten, SSB, Telenor and Gjensidige are interested in the prediction of certain behaviours of their customers and service users, predicting churn or fraud activities. In the health area, the availability of real time monitoring of patients and healthcare insti-tutions allows completely new screening protocols and treatment monitoring, real time prevention and increased safety. For ABB and DNV GL high dimensional times series are generated by sensors monitoring a ship, with the purpose of predicting operational drifts or failures and redesigning inspection and maintenance protocols. The objective is to predict the dynamics, the future performance and the next events. Importantly, real time monitoring of such transient behaviour and a causal understanding of the factors which affect the process, allow optimal interventions and prevention. While the concrete objectives are diverse, we exploit very clear parallels:

- systems operate in a transient phase, out of equilibrium and exposed to external forcing;
- in some cases, there are many time series which are very long and with high frequency; in other cases, short and with more irregular measurements;
- complex dependence structure between time series;
- unknown or complex causes of abnormal behaviour;
- possibilities to intervene to retain control.

BigInsight develops new statistical methodology that allow our partners to produce new and more precise predictions in unstable situations, in order to make the right decisions and interventions.



ORGANISATION

Board in 2018

Marcus Zackrisson, Skatteetaten, chairman
Andree Underthun, ABB
Karl Aksel Festø, DNB
Bjørn Johan Vartdal, DNV GL
Birgitte F. De Blasio, Folkehelseinstituttet
Erlend Willand-Evensen, Gjensidige
Ellen Charlotte Stavseth Paaske, Hydro
Ulf Andersen, NAV
Lars Holden, Norsk Regnesentral
André Teigland, Norsk Regnesentral
Peder Heyerdahl Utne, Oslo University Hospital
Anders Holmberg, SSB
Astrid Undheim, Telenor
Bård Støve, University of Bergen
Nadia Slavila Larsen, University of Oslo

Observer: Terje Strand, Research Council of Norway
The board had 2 meetings in 2018.
All partners are represented in the Board.

Legal organisation

BigInsight is hosted by NR.
Legal and administrative responsible:
Managing director Lars Holden

Center Leader

Prof. Arnoldo Frigessi, UiO Director

Co-Directors

Ass. Research Director Kjersti Aas, NR
Prof. Ingrid Glad, UiO
Ass. Prof. Ingrid Hobæk Haff, UiO
Ass. Research Director Anders Løland, NR
Research Director André Teigland, NR

Principal Investigators

Kjersti Aas, NR
Magne Aldrin, NR
Arnoldo Frigessi, UiO
Ingrid Glad, UiO
Clara Cecilie Günther, NR
Ingrid Hobæk Haff, UiO
Alex Lenkoski, NR
Anders Løland, NR
Carlo Mannino, UiO
Magne Thoresen, UiO

Administrative Coordinator

Unni Adele Raste, NR

Scientific Advisory Committee (SAC)

Prof. Idris Eckley, Lancaster Univ., UK
Prof. Samuel Kaski, Univ. Helsinki, Finland
Prof. Geoff Nicholls, Univ. Oxford, UK
Prof. Marina Vannucci, Rice Univ., Houston, USA
Senior Lecturer Veronica Vinciotti, Brunel Univ. of London, UK

BigInsight


BOARD
NR´S RESPONSIBLE PERSON
SCIENTIFIC ADVISORY COMMITEE
DIRECTOR
CO-DIRECTORS
PRINCIPAL INVESTIGATORS AND CO-PRINCIPAL INVESTIGATORS
INNOVATION OBJECTIVES
Personalised marketing
Personalised health and patient safety
Personalised fraud detection
Sensor systems
Forecasting power systems
AI – Explanation & Law

PARTNERS

Partners

- Norsk Regnesentral (host institute) (NR)
- University of Oslo (UiO)
- University of Bergen (UiB)
- ABB
- DNB
- DNV-GL
- Gjensidige
- Hydro
- Telenor
- NAV (Norwegian Labour and Welfare Administration)
- SSB (Statistics Norway)
- Skatteetaten (Norwegian Tax Administration)
- OUS (Oslo University Hospital)
- Folkehelseinstituttet (Norwegian Institute of Public Health, NIPH)
- Kreftregisteret (Cancer Registry of Norway)

Cooperation between the partners of BigInsight

There have been two board meetings in 2018, where all partners are represented. In addition to close cooperation with the researchers at NR and the universities, there have been several meetings within the separate Innovation Objectives where partners have met and exchanged ideas.

The seminar series "Opening the black box" have brought all partners closer together and has resulted in more bilateral partner-to-partner cooperation across the Innovation Objectives.

In October the annual BigInsight Day was held at the premises of DNB. In this successful event a wide selection of our projects, ideas and results were presented and discussed among representatives from the different partners and project members.

RESEARCH STRATEGY

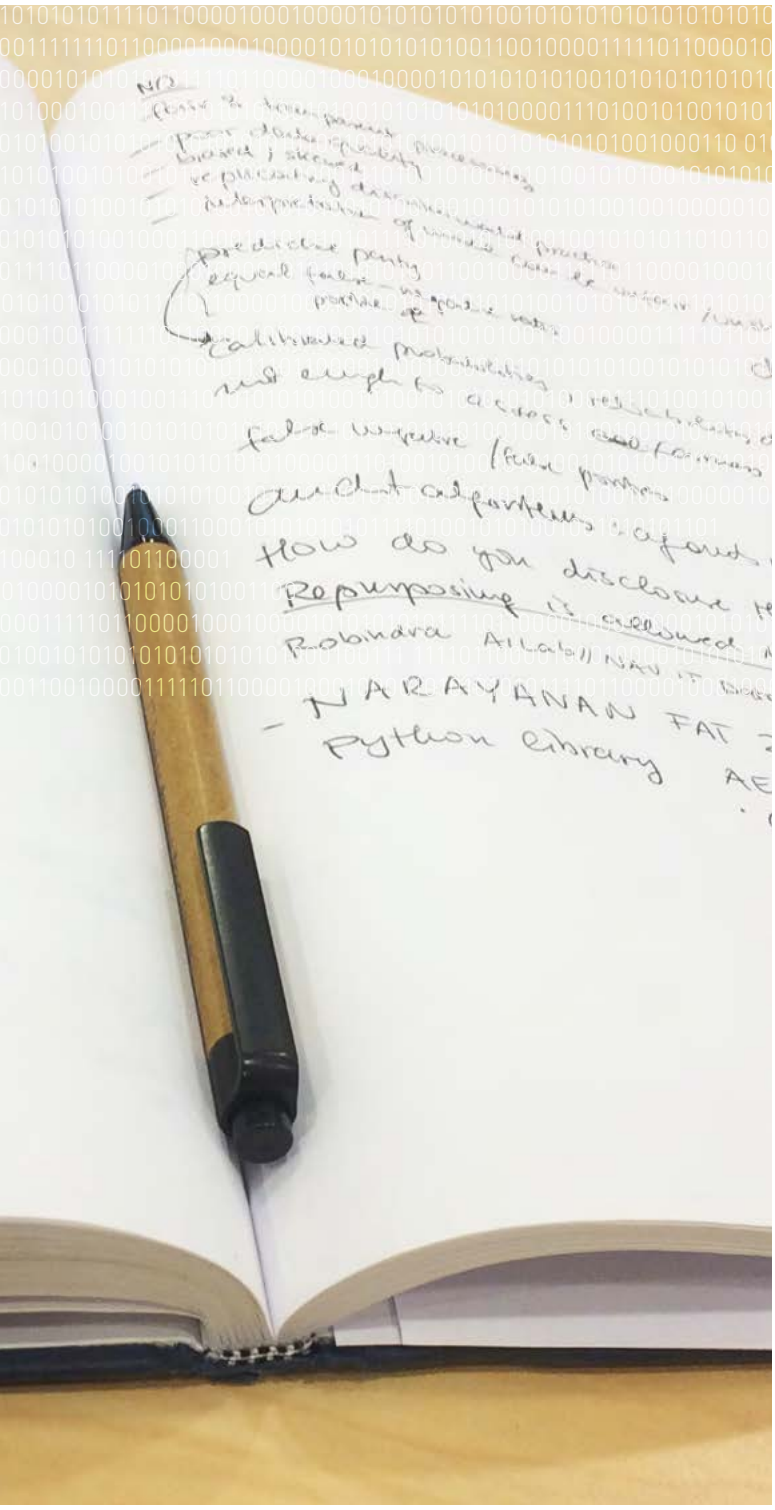
BigInsight's research is organized in six innovation objectives. Five innovation objectives (IOs) are centered on a concrete innovation area: marketing, health, fraud, sensor, power. The last IO is focusing on explanability of AI and data privacy.

Each IO has specific innovation aims related to outstanding open problems, which we believe can specifically be solved with new statistical, mathematical and machine learning methodologies. Our research projects deliver methods and tools for their solution. We aim to new, interesting and surprising solutions, which take the field and our partners ahead in their innovation strategy. Final transfer to partners' operations will happen both within and on the side of BigInsight.

INNOVATION OBJECTIVES	 Personalised marketing	 Personalised health and patient safety	 Personalised fraud detection	 Sensor systems	 Forecasting power systems	 AI - Explanation & law
INNOVATION PARTNERS	DNB Gjensidige NAV Skatteetaten Telenor SSB	DNV-GL Kreftregisteret OUS Telenor	DNB Gjensidige Skatteetaten	ABB DNV-GL SSB	DNV-GL Hydro Energy SSB	all partners
RESEARCH PARTNERS	NR UiO NIPH UiB	UiO OUS NR NIPH	NR UiO UiB	NR UiO	NR UiO	NR UiO
PRINCIPAL INVESTIGATORS						
Principal Investigators: co-Principal Investigators:	Kjersti Aas Arnoldo Frigessi	Magne Thoresen Clara Cecilie Günther	Anders Løland Ingrid Hobæk Haff	Ingrid Glad Magne Aldrin	Alex Lenkoski Carlo Mannino	Anders Løland Arnoldo Frigessi

METHODS

We solve innovation challenges of our partners by developing solutions, which are based on new statistical, mathematical, and machine learning methods.



Our recent methodological results include:

- Integrative analyses of complex multiple data sources, including integrative clustering and methods to investigate coordinated architectures across clusters in various data sets.
- High dimensional penalised regression, also assuming monotonicity, and with measurement error in covariates, with applications to genomics.
- Bayesian hierarchical models, including monotone multiple regression and cancer drug synergy prediction, with applications to insurance, drug screening, recommender systems and mortality data.
- Inference and prediction in multiscale models of stochastic differential equations in bio-mathematical models.
- Models for infectious disease processes, including the effect of urbanisation on epidemics and the use of mobility cellular phone data to describe the contact process.
- General methods to describe uncertainty in predictions, with applications to power market.
- Pair copula constructions for structure learning in Bayesian Networks for financial data.
- Models for the forming of social networks and inference from data in time.
- Anomaly detection algorithms with sequential build-up of anomaly evidence.

SCIENTIFIC ACTIVITIES

BigInsight researchers are working on several research projects, motivated by our partners' innovation needs. They cross disciplines and industrial sectors and challenge the available state-of-the-art. New methodology is developed and tested on specific innovation cases and data from the partners.

Each IO has a research team, with members from the relevant innovation and research partners. This includes senior and experienced staff as well as junior staff, postdocs, PhD and master students and international collaborators.
The following are highlighted results from 2018:

A more efficient market for users and customers

Written by Elin Ruhlin Gjuvsland, Senior Communication Consultant, Norwegian Computing Center

What do the defaulted DNB corporate and mortgage customers, the viral spread of Telenor's products and the click pattern in Gjensidige and Skatteetaten web data have in common?

The answer is methodological developments in BigInsight. Kjersti Aas, co-director of BigInsight, is responsible for the projects under the umbrella of personalised marketing. Although each project is different, they share several common denominators.

- In their own way, each project is about better risk management and optimization of marketing, targeting each user or customer as an individual unit, says Kjersti.

Satisfied partners

DNB currently has 25% of Norwegian mortgage customers. With today's methods of approving loan applications, little consideration is given to whether the applicant is smart in personal finances. Although people may have the same income, the risk of default can be very different.

BigInsight has looked at how machine learning and account movement data can be used to distinguish between good and bad customers. Kjersti and her team found, among other things, that large fluctuations are negative, while good customers have more regular patterns in time. At DNB, the deep learning algorithm that Kjersti's team

has developed is now ready for use. It will allow to decide whether DNB gives customers loans in a more automatic and rapid fashion, leaving only complicated cases to further manual check. The first tests of the method indicate a huge saving of time and work and an excellent prediction power. More customers appear to be able to obtain a loan right away.

- Such success stories are, of course, the best ones. In the end, our partners are happy, we can see the direct impact of our research, our PhD students make an important experience and we publish good science on central journals. It ticks all the boxes, says Kjersti, smiling.

Networks in businesses

Do individuals with key roles in defaulted businesses increase the chance of other corporate defaults if they are also related to those businesses?

- We are in the process of understanding these links and are excited about what we will discover. If a link exists, it narrows the opportunities for key people who are constantly driving businesses down, Kjersti says. Studying networks is a very central methodological challenge in many areas of BigInsight.

Click Patterns

The way we click around online says something about what we are going to choose the next time we visit a website. This allows personalisation of web pages.

Together with DNB, Gjensidige and Skatteetaten, BigInsight has gone deeper into the information collected every time an individual visits a given website. Although the data from these major players are of course not linked, the methodological issues they face are very similar.

- We look at whether an individual's online patterns can say anything about the likelihood of their next choice, Kjersti explains. We also look at whether the website organisation should be improved. For example, we investigate if the length of time an individual spends on a page is associated to how complicated the page is.

Viral products

Is it possible to predict early whether a product will be a success or a flop by analysing how a product spreads in a social network of potential users and customers?

- We know viral marketing, ie the natural and self-triggered spread of information on social media, is highly efficient, says Kjersti. We want to predict such a viral behaviour, once only a very short period of sales is observed in the market. We used a case from the period right after the launch of a recent Telenor´s app to build a model to simulate spread on the network of Telenor's mobile phone users, further ahead in time. The results are very promising, and with small changes, the same model can be used for many other types of products, when a network of interactions in the market is available.

Recommender Systems

Website recommendation systems that give suggestions about other products of the same genre, still have a way to go, says Kjersti. Therefore, BigInsight has launched a project to improve these recommendation systems. To test their new methodology and algorithms, BigInsight used data from NRK. Arnoldo Frigessi, who leads a team working on preference learning and recommender systems, believes their method has two clear advantages:

- First our method appears to make more diverse suggestions, compared to state-of-the-art recommenders. In this way the catalogue of items to be recommended is exploited better, in its full breadth, and the recommendations appear to be more personalised. Second, our statistical approach estimates an uncertainty level for each suggested item. This is important, because recommendations should be probably simply avoided when they are too uncertain. This is thanks to our Bayesian approach!

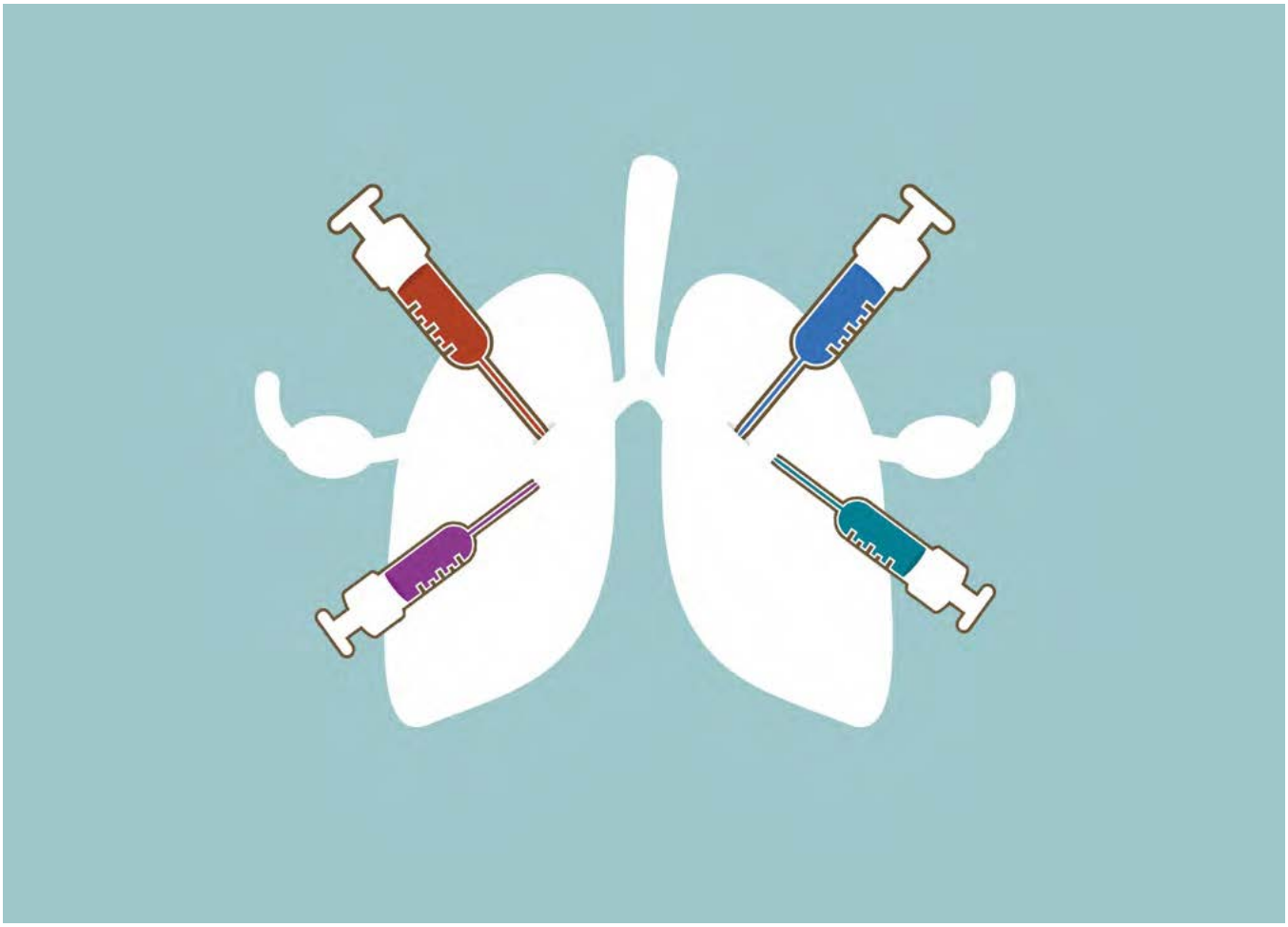
Sometimes it does not work

For Kjersti and the personalised marketing teams, it is a thrill to work when the projects end in success. However, there are also good lessons to be learned when projects go wrong.

- We usually end up with important knowledge also when projects don't work out. In the very least, we have learned when to end projects, in time before we have spent too many resources on them, says Kjersti.

If a project fails, it is most usually due to insufficient and too noisy data for the question to be answered and not because of lack of methods to analyse them.

- You don´t always get good results when you have large amounts of data. The main reason is that data does not always contain the information needed to solve the customer's problem.



In the future, simulations on the computer may make it possible for oncologists to choose the drug that is suitable for each patient.

Simulating your cancer treatment on a computer

In ten years, computers will be able to propose the most suitable cancer treatment for you. The idea is to simulate how all possible combinations of existing cancer treatments will work on your particular tumour.

Text: Yngve Vogt, Apollon

Each year, 10,000 Norwegians die of cancer. Researchers at the University of Oslo are now developing a computer program that can help oncologists find the best personalised treatment for each patient. The hope is to be able to cure far more patients.

One of the most common cancer diseases with poor prognoses is lung cancer. Each year, 3,000 are diagnosed with this sinister disease. 2,200 die of it.

In fact, lung cancer takes as many years of life as breast cancer, prostate cancer, and intestinal cancer combined.

"Although treatment has improved, we are still lacking tools to be able to decide the best treatment for the individual", says Åslaug Helland, who is both a professor at UiO and clinician for lung cancer patients at Oslo University Hospital, Radium Hospital.

Some of her patients start treatment with chemotherapy, others with immunotherapy. Immunotherapy is a modern treatment that stimulates the body's own immune system to fight the cancer. Although many people get good help from immunotherapy, it only works for half of the patients.

"The problem is that we don't know who benefits from immunotherapy. If the drug does not work after two or three rounds, we will try something different, such as the classic chemotherapies, but by then we will have already squandered time that could have been spent on other treatment. We therefore need a system that says what is the best and most effective treatment for each patient – and with the fewest possible side effects", Åslaug Helland points out.

Arnoldo Frigessi is the main person behind the new idea of making tens of thousands of virtual copies of the individual cancer patient and then testing all the treatments in a simulation model to find the most optimal treatment. Photo: Ola Sæther

This is precisely the great idea of Arnoldo Frigessi, professor of statistics at the University of Oslo, who leads both the Oslo Centre for Biostatistics and Epidemiology and BigInsight, a centre for research-based innovation.

Simulates cancer treatment

His idea is to use mathematical and statistical methods to develop a computer program that will propose the best cancer treatment. Arnoldo Frigessi is an expert in describing biological processes with mathematical models. He works closely with everyone from oncologists and pathologists to molecular biologists, statisticians and mathematicians.

"Today's treatment does not help three out of ten cancer patients. The cancer cells can also become resistant to the drugs they receive. The drugs then need to be replaced. Might one of the solutions be to give the patient several cancer drugs at the same time?" asks Arnoldo Frigessi.

This is far more challenging than one would think. Imagine that your oncologist has the possibility to choose between 300 different drugs for a particular cancer disease. If the doctor is to choose the best combination of two different drugs from this selection, he has as many as 45,000 different treatment choices.

"This multiplies the possibilities to the point where it becomes impossible for the doctor to figure out which combination of drugs works best".

It is not possible to test all the possible combinations of drugs with traditional clinical trials.

In classic clinical trials, two groups of patients are compared. One is given drug A, the other drug B.

However, the doctor now wants to provide personalised medicine to a patient group that consists of just one patient.

This makes it impossible to carry out classic trials of different types of treatments.

Instead, Arnoldo Frigessi's solution is to make hundreds of thousands of virtual copies of the patient and test all the treatments on a simulation model on the computer to find the best treatment.

Can change treatment

Lung cancer expert Åslaug Helland believes that such a simulation model can change the future of cancer treatment. "We strongly believe that it can be possible to attack the

Over two thousand Norwegians die of lung cancer each year. It is difficult to find the right treatment. Now, Chief Physician Åslaug Helland hopes that future simulation of cancer treatment can make it easier for doctors to find the right treatment, tailored to the individual patient. Photo: Yngve Vogt

cancer from different fronts at the same time. Then we need to know which combinations of drugs work best in each individual case", says Åslaug Helland.

The combination of options increases even more by testing different doses and the order in which the drugs are taken.

"There are endless possibilities here, and there's a lot we don't know. The simulation model can help us determine which combination is the best treatment for the individual patient, rather than spending time on treatments that do not help", Åslaug Helland points out.

This simulation model is a so-called multi-scale model, with mathematical models of what is taking place from the molecular level up to the size of the cancer tumour.

"We collect all possible data from the patient and can calculate how the treatment affects the chemical and physical reactions in millions of different cells. We can then simulate how the cancer would have developed with all the different therapies. Machines work fast. Our goal is that we might be able to test a hundred thousand different treatments over the course of one night and then say which ten drugs work best", points out Arnoldo Frigessi.

To accomplish this, scientists need to mathematically describe how the lump of cancer functions, how many cells it consists of, how dense the cancer cells are and where the blood vessels are. They have also described the energy balance and the chemical and physical reactions in the cells.

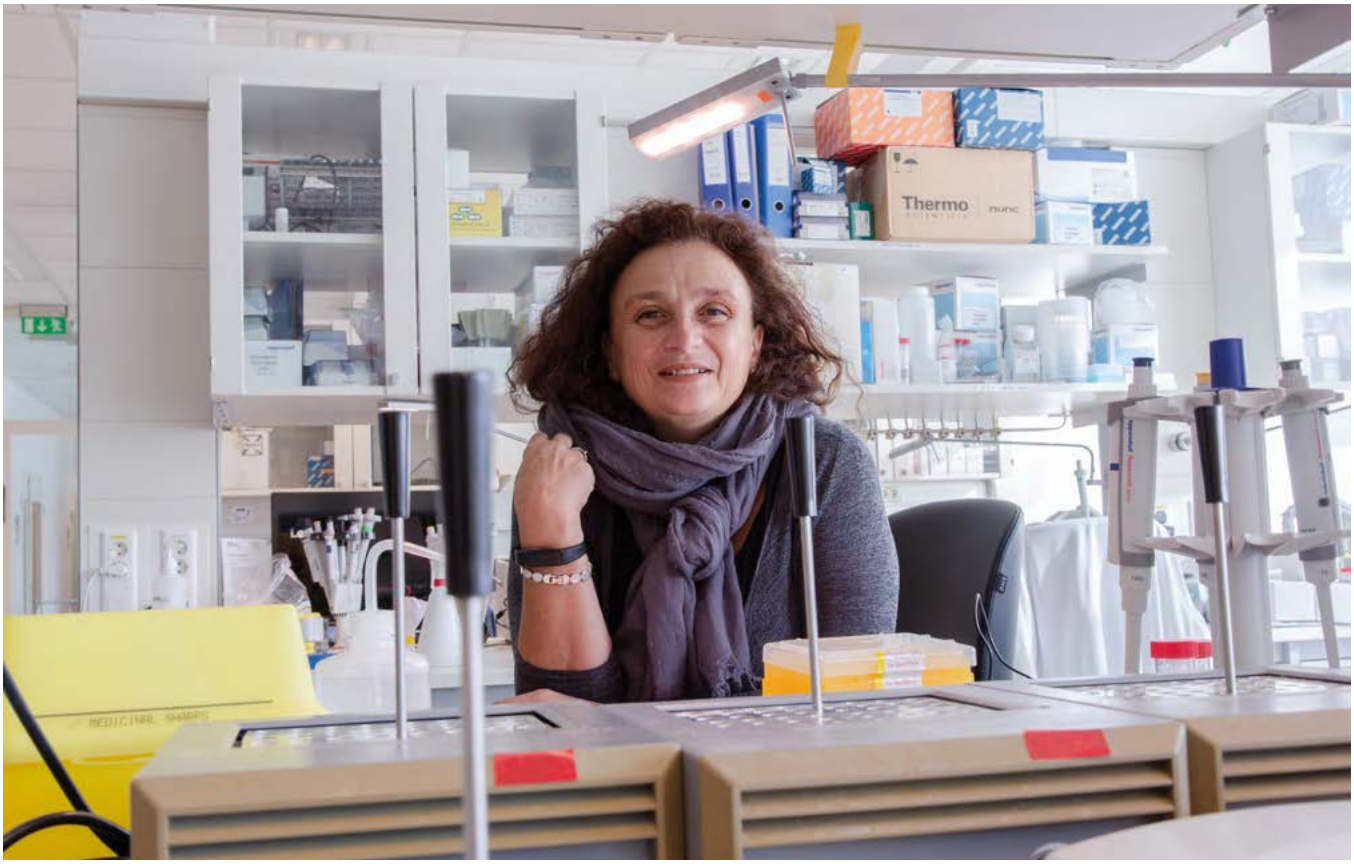
Their data includes genomic biomarkers, histopathology and magnetic resonance images.

"We can then use the mathematical model to describe the condition of the cancer cells for the next twelve weeks, while bombarding them with various medications. Some drugs can only kill the cell at the moment it splits into two. We describe how each cell lives, develops, splits, and dies. We also need to calculate the amount of medicine needed to kill the cells and the probability of this happening".

Oxygen to the cancer cells

The researchers also model how the cells talk to each other, how the blood vessels grow and die and how the oxygen is distributed in the tissue.

The blood carries oxygen to the cancer cells. If the supply of oxygen stops, the tumour may die.



Vessela Kristensen notes that the simulation model can improve current cancer treatment so that more survive the disease. Photo: Yngve Vogt

"The cells need oxygen to breathe. If the cells don't get enough oxygen, they're in trouble. The mathematical model explains how the cells are doing depending on the amount of oxygen", says Arnoldo Frigessi.

The simulation model is very complicated. They have to build it up piece by piece. An essential part is precisely to simulate what is needed to reduce the supply of blood to the tumour.

Some of the simulations show that the cancer flares up again.

"It may be because the drug was given too seldom or at the wrong time".

Cancer needs a lot of blood to grow. To get the blood they need, cancer cells produce a special protein that gives notice where the blood vessels should grow. Their model must therefore also take into account how the amount of this particular protein changes. A delicate balance is at work here.

"If the blood vessels are destroyed, the chemotherapy will not reach the cancer cells".

The drugs follow the blood vessels and must leave the vessels where the cancer cells are located. The program describes how much medication is leaching from the blood vessels and when it occurs.

"This varies from patient to patient and must therefore be determined before the simulation can be run".

The procedure is simple. The patient takes a contrast agent. Then it is possible to study how quickly the contrast agent leaches out of the blood vessels on an MRI scan.

Unfortunately, the information is well hidden in the enormous amounts of data from the MRI scan. Arnoldo Frigessi must therefore develop statistical methods in order to determine how fast the blood vessels of the individual leach drugs.

Numerical paradise

The simulation model is bursting with stochastic models, which describe random events, multi-scale models, which provide a mathematical description of how the cancer develops in different dimensions, and last but not least: partial differential equations describing developments in time and space.

The simulations must be run hundreds of thousands of times. It is then important to optimise the program so that they run as fast as possible.

This is where the numericians enter the scene. Numerical analysts are experts in solving mathematical problems with the computer. Here, Arnoldo Frigessi has

roped in Marie Rognes, chief researcher at the Simula Research Laboratory.

With her expertise, the equations can be solved many hundred times faster. It will then be possible to simulate much larger parts of the tumour without losing time.

"We simulate how the tumour develops every half hour. Numerically speaking, it is impossible to do it less often. The calculations will then be inaccurate. A number of calculations must be made for each unit of time. And the simulations must be conducted for all the different proposals for each cancer treatment", says Marie Rognes.

She has taken two big steps to optimise the program.

One is to create a code that guesses the right solution far faster. The second thing that she will eventually work on is to program the code so that the simulations can be run in parallel on a high-performance computer simultaneously by thousands of data processors.

Tested on breast cancer

The simulation model has been initially tested on patients with breast cancer. There are a number of different types of breast cancer.

3,500 contract breast cancer each year. Nine out of ten survive.

Researchers have received detailed data from Ullevål Hospital/Radium Hospital on how four patients who had already been treated for breast cancer fared.

"The simulation program reached the same results about how the cancer developed".

This means that their simulations matched the reality.

In their simulations, they have currently only modelled the effect on two types of chemotherapy.

"We must include all the drugs that are used. They all work differently", says Frigessi.

Breast cancer can be divided into two main categories: Cancer cells with and without oestrogen receptors. Cells with oestrogen receptors are the easiest to stop by blocking the receptor. The other group is more difficult. Chemotherapy works well on those cells that divide quickly, but unfortunately some patients suffer a relapse.

Clinicians have guidelines on how they can treat breast cancer, depending on how the cancer cells look and what kind of receptors they have.

"These guidelines build on many years of experience and knowledge from clinical trials. There is a plethora of treatment options and possibilities. Before we go and throw ourselves at all the medical combinations outside the guidelines, there are still niches within these that we can study further", says Professor Vessela Kristensen at the University of Oslo. She is one of the country's foremost researchers in breast cancer.

She says it's important to knock out all the cancer cells.

Chief Physician Olav Engebråten, who is an expert on breast cancer, is sceptical about whether simulations can improve cancer treatment in the short term, but still believes that the idea is exciting over the long term. Photo: Yngve Vogt

It's not enough to knock out the lion's share.

"The few remaining cancer cells are the dangerous ones. The point is therefore to have a mathematical model for those patients whose treatment is not working. Although most people survive breast cancer, which is the most frequent form of cancer in women, the small percentage that suffer a relapse still totals many women", Vessela Kristensen points out.

Sceptical

Chief physician Olav Engebråten, who is an expert in the project, is sceptical about whether the simulations can improve treatment in the short term. He is a clinician and one of the foremost experts in treating breast cancer at Oslo University Hospital.

"The idea is exciting. It may be interesting in the long term, but the problem is the wide variation that cancer diseases pose per se. The trademark of cancer is variation and instability. The systems must be good enough to pick out the medications that work or do not work based on the characteristics of the tumour", Olav Engebråten points out.

Some patient groups receive targeted treatment. Others do not. Among those who do not receive targeted treatment, there are seven to eight different standard treatments.

"Some standard combinations are well-proven", says Olav Engebråten.

He is particularly sceptical about having to deal with a

computer program that will suggest an optimum combination of drugs that have not been tested before.

"Combining and adjusting drugs is very challenging. Many of them work on the bone marrow and other tissues. In the clinic, we usually do not combine drugs until the effect is well tested. When we calculate a patient's drug dosage, we need to know if the patient can withstand the treatment. The normal tissue must survive", he says, adding:

"We will always experience that patients die from the disease. We can't cure everyone, but it's extra important that we do what we can to prevent patients from dying from the treatment itself. We can't throw precaution overboard".

He is also sceptical about the current simulation model of the blood vessels in the tumour.

"This model is a simplification of the reality. All of the assumptions are not enough. We also need more data", notes Engebråten.

However, he adds that in the long term the simulations may be interesting for those patients who are not benefiting from current treatment.

The survival rate for patients with breast cancer is up to 90 per cent. However, the prognoses are poorer for those with extensive illness, such as when it spreads to the lymph nodes.

"The poorer the prognosis for survival, the greater the potential of the simulation model. We might then be able to save more lives in the long term. It would therefore have been fantastic to have such a simulation tool that could help us to choose the right treatment for such patients", notes Engebråten.

Arnoldo Frigessi appreciates the scepticism of the clinicians.

"I'm glad that the clinicians are sceptical. It helps us to improve the models so that they can become more realistic. I think one day that simulations will be part of clinical practice, but we are not there yet", says Frigessi.

Åslaug Helland is not as critical as Olav Engebråten. She thinks it might be interesting to simulate blood vessels.

"Although it is natural to think that we kill a tumour by taking away its blood supply, the results of this strategy have varied. We can still use the simulations to show who can benefit from this treatment", Åslaug Helland hopes.

Vessela Kristensen notes that certain treatments destroy the formation of blood vessels.

"Simulation of blood vessels is still only one of the many things that are part of the simulation model. The number of blood vessels is easy to measure. We are therefore trying to simulate the cancer-fighting effects of destroying blood vessels".

She believes the model can show a correlation with the blood vessels not previously known about, adding:

"Currently, we only simulate a small part of the reality. The simulation model can still change future cancer treatment. Medical treatment is based on experience. This experience takes a long time to build up. Sometimes it doesn't work. The simulation model can accelerate this process", Vessela Kristensen believes.

Her entire point is briefly summarised:

"Now we don't have to wait for generations to die to accumulate enough experience. The mathematical models can find the essence of all the experiences and come to faster conclusions. We therefore save both time and lives with these models. They can improve current practice", Vessela Kristensen believes.

Mouse model

When the simulation program has made its proposals, the oncologists can double check the treatment in the laboratory before the patient is medicated.

"One possibility is to take pieces of the tumour in different vessels and test whether the cancer cells die with the various combinations of drugs", says lung cancer specialist Åslaug Helland.

Another possibility is to test the proposed treatment on mice or in zebra fish. A few cells are then taken from the tumour and inserted into the test animals, which receive individual treatment.

In the United States, clinics are already testing the tumour in mice and trying to treat mice with different types of treatments. The method nevertheless has its weaknesses:

"Not all tumours grow in mice, the experiments use a lot of mice, and the method is very expensive", says Vessela Kristensen.

Probability

Arnoldo Frigessi emphasises that the simulation model will not be able to provide a hundred per cent certain result.

"However, the doctor will receive an answer from the program whether the result is 80 or 40 per cent certain. It makes a difference. The program must be able to say that the result is uncertain because some parameters, such as the amount of medication leaching from the blood vessels, are uncertain. I therefore believe that it is important to quantify the uncertainty. The program must also be able to say whether the treatment of a patient will be easy or difficult", Frigessi points out.

Norwegian Medicines Agency

Ultimately, the clinicians must have confidence in the simulation program.

"No doctors dare use the recommended combination treatments when no one has done it before. The doctors must have confidence in the algorithm. My question is how

to enable doctors to put their trust in these algorithms".

His idea is that the algorithms in the simulation program need to be approved by the authorities before they are used.

"Because the recommended combination of drugs is tailored to the individual patient, the Medicines Agency must approve the algorithms for the therapy and not the therapy itself.

Perhaps our simulation program can be approved for every single disease, such as breast cancer and lung cancer", Frigessi hopes.

Uncertain future for fraudsters

Fraudsters who cheat on their taxes, launder money or con insurance companies are facing uncertain times. New statistical methods are increasing the likelihood that they will be caught.

Text: Yngve Vogt, Apollon

The underground economy in Norway is flourishing. Each year, at least NOK 150 billion in direct and indirect taxes is lost to evasion.

The Norwegian Tax Administration, Gjensidige Forsikring and Den norske Bank (DnB) have now joined forces with researchers at the University of Oslo and the Norwegian Computing Center to outmanoeuvre financial criminals. Their project has been given the resounding name "Personalised Fraud Detection".

Cheats

DnB needs better methods to hunt down money launderers, an activity they are legally required to report.

Gjensidige Forsikring wants to expose those who report damage that is more serious than in reality, and those who claim excessive values for stolen items.

The Tax Administration wants to adopt new methods to combat the underground economy. The project is good news for all those who do not cheat on their taxes: "Those who follow the regulations should not lose out against those who do not follow the regulations".

Thanks to new digital capabilities, the three actors will have access to far more information than in the past. This offers tremendous opportunities.

Cheating on VAT

One of the Tax Administration's main issues is to find new ways to expose those who cheat on reporting value-added tax (VAT).

Each year, the Tax Administration receives 1.6 million VAT returns. Many of them contain errors. The errors are caused when filers deliberately overreport expenses, underreport revenues, misunderstand the regulations or make incorrect entries.

It is impossible to manually check all VAT returns.

"The Tax Administration must therefore have an automated system to screen out prospective tax cheats. The case must then be forwarded to a case manager for more a thorough assessment", says Ingrid Hobæk Haff, associate professor at the Department of Mathematics at the University of Oslo.

Each year, 34,000 new VAT-liable companies are established. The Tax Administration also wants to create a statistical model that can indicate the risk of who is going to cheat already from day one.

Related fraud

Although money laundering differs from insurance fraud and tax fraud, there are some similarities. The similarities will be used to develop completely new statistical methods that will uncover the swindlers.

The three collaborators are not going to exchange each other's datasets, but by taking on different datasets with related issues, the statisticians can make strides.

Today's methods do not work well enough. Ingrid Hobæk Haff is therefore developing new methods for exposing fraud.

"Each time a new case comes in, we want to calculate the probability of fraud".

The project is part of BigInsight, a centre for research-based innovation at UiO. The centre's speciality is developing all-new statistical methods to uncover irregular patterns in enormous amounts of data.

The goal of the researchers is to calculate the probability that something is wrong and that the case should

MONEY LAUNDERING: New statistical methods can outsmart money launderers. Photo: Ola Sæther

undergo further scrutiny.

Most attempts at fraud are never checked. Even if a case is worthy of examination, it is still not certain that it pays to check it out. The computer program will therefore simultaneously estimate the expected gain from examining the case.

Exposing the swindlers

The new statistical methods are tested on datasets whose content has already been checked. That means that the statisticians know in advance whether or not someone is a swindler.

"We can then check our results against the answer. The downside is that we don't get to check our program on the cases that have not already been checked. The cases that were checked might have had something suspicious about them in the first place, so they weren't selected at random".

The control dataset can therefore be skewed because the cases that have been manually selected for control have already been chosen according to certain criteria.

Ingrid Hobæk Haff therefore wonders whether there is information in the rest of the dataset that complements the information they already have, and whether it is possible to exploit this information in some way.

"We must then distinguish between 'Yes, we know the answer,' and 'No, we don't know the answer'. The whole point is to find the data that stands out".

The data that stands out doesn't necessarily have to be fraudulent.

And although the new statistical method can expose more fraud, it is not certain they will be able to find everyone. That doesn't matter that much.

"If we make the method just a little better, it can still lead to huge gains".

The method must also take into account that fraudsters change the way they cheat.

"If we expose fraud, the fraudsters will try other methods instead. We must therefore continually develop new methods".

Mathematically tricky

One of the many mathematical intricacies is unbalanced datasets. Fraud cases constitute only a small percentage of all cases. Since it is difficult to pick them out, there is a

STATISTICIAN: Ingrid Hobæk Haff is developing new statistical methods that will make it possible to find cases with a high probability of fraud. Photo: Yngve Vogt

FACTS
Financial crime

- The underground economy is flourishing in Norway. Den norske Bank, the Norwegian Tax Administration and Gjensidige Forsikring are now stepping up the fight against financial crime.
- The three giants are working with statisticians at the University of Oslo and the Norwegian Computing Center.
- Together, they are developing new statistical methods to outsmart financial criminals.
- The goal is to find the cases where the probability of fraud is high.

risk that the case managers will be told to check out many law-abiding cases in relation to fraudulent cases.

"We only want to find the cases where the probability of fraud is high".

Another tricky problem for the mathematicians is not only the large quantity of data, but also the many dependencies between the different data.

"Because multiple variables in a dataset can provide the same information, many of the statistical methods start to struggle".

And as if this is not enough.

"We have a lot of information with different statistical properties. Some of the information may have all possible values and some may have only a few values. This combination can be mathematically difficult".

Business secret

Norway is not the first country in the world to conduct research on how to detect fraud. Ingrid Hobæk Haff has therefore checked what other researchers have done.

"We have tested their methods, but they don't solve our problems the way we would like".

However, this does not mean that no one in the world has tried to crack the fraud code.

"I would certainly assume that big insurance companies and banks out in the big world have solved the problem with their own statistical methods. They don't publish their solutions. They therefore remain well-guarded secrets", says Ingrid Hobæk Haff.

SOUND THE ALARM ON TIME: "New statistical methods should make it possible to sound the alarm BEFORE faults on large ships happen", say Morten Stakkeland and Ingrid Glad. Photo: Yngve Vogt

Avoids ship disasters with advanced statistics

The chance of major ship disasters at sea can be reduced by statistical methods. The trick is to interpret the large amounts of data streaming in from the many sensors in the ship, making it possible to sound the alarm on time.

Text: Yngve Vogt, Apollon

Companies can save millions of kroner by using advanced statistical methods to extract information from large amounts of collected data.

One company that has adopted this new way of thinking is the international industrial group ABB. They work closely with statisticians at the University of Oslo to enable shipping companies to save millions on more optimal operations and improve ship safety.

Ships of today are checked at regular intervals to ensure that they maintain their condition. The ships must be docked to do this. Such breaks can be quite costly.

"Instead of sending inspectors on board ships, we can rather use statistical methods to monitor all the data for changes, such as when the engine is about to overheat or break down. This automatic data monitoring should be able to provide warnings about incidents before they happen", says Morten Stakkeland, who is both a project engineer with ABB and adjunct associate professor of statistics at UiO.

If the captain is notified a few hours before the engine breaks down, it is possible to repair it before the ship shuts down. Imagine the disaster if an engine fails just when a several hundred-metre-long tanker is approaching an oil terminal.

The main mantra of the statisticians is to extract and interpret the enormous amount of information that can be collected from large ships.

"The point is to be able to find a pattern in these reports, so that the alarm sounds when something approaches a crisis", says Ingrid Glad, professor at the Department of Mathematics at the University of Oslo. She is part of the research group Big Insight, a centre for research-driven innovation. They are working to develop and commercialise statistical methods.

Sensors

The data comes from the huge amount of sensors on board. A single ship can have several thousand sensors. Some sensors transmit data every ten seconds. Other sensors transmit data as often as every millisecond. As one can imagine, this amounts to a huge amount of data over the course of a year.

Sensors are not the only source of data. The large amount of equipment on board also communicates with each other.

On large tankers, the instruments send a series of cryptic reports to each other. None of these reports are standardised.

"Several gigabytes of information are gathered in a short period of time".

The huge amounts of data are continuously stored on the ship or sent to control centres on shore.

The point is to find something that stands out in the data streams.

"We can then build a statistical model that shows the normal condition and where we can find deviations. We might find deviations in only one sensor, or perhaps all the data from all the sensors are within the accepted range, but the correlations between the data are so special that they still trigger an alarm", points out Erik Vanem, senior researcher at DNV-GL.

Checking the hull

Cracks in the ship's hull is one of the feared scenarios. The struggle against big waves in rough seas can in the worst case break the hull apart.

Hundreds of sensors can be placed inside the hull to monitor the risk of material failure.

DNV-GL, formerly Veritas, looks for tension in the hull. They specialise in ship safety and work on making sea transport safer.

"We can measure the tension in the hull and use this to say something about the strength of the ship and the risk of fatigue. In addition to inspecting ships on a regular basis, we can equip critical components with sensors to provide continuous monitoring. We can then catch things that we can't catch with spot checks, and use the additional knowledge to plan when physical inspection is needed, rather than having to inspect the ship at fixed intervals", says Erik Vanem.

The sensor data is also linked with weather data to calculate the extent of the strain on the hull over time.

"If a ship has taken a lot of beating, we can pay extra attention", says Erik Vanem.

Battery check

Even something as seemingly simple as a battery check can quickly save shipping companies large sums of money. Electric ships with large batteries must be taken out of operation a whole day for the annual service.

"This is lost uptime. We look at how it is possible to use statistics to evaluate the batteries. This will save shipowners a lot of money", says Morten Stakkeland.

One of the big points is maintenance only when needed, rather than at fixed intervals.

"Shipowners could potentially save millions of kroner per ship", Stakkeland points out.

He believes that data monitoring may be the first step towards driverless ships.

Advanced statistics

The big challenge is being able to interpret the enormous amounts of data and sound the alarm when necessary, but not otherwise.

"The ultimate goal is to be able to report that the engine is going to fail before it happens", says Ingrid Glad.

She compares this to looking for something abnormal in the woods without knowing what it is.

"If you go mushrooming, you know exactly what you're looking for. However, we are looking for something we have never seen in the woods before. That's much harder".

The analysis must also take place in real time. This means that the statisticians need to create a system that detects anomalies before the abnormal situation leads to faults or damages. At the same time, they must reduce the likelihood of false alarms. It's a mathematical challenge.

Assembling and mining the largest dataset of the Nordic power market produces competitive advantages

Written by Elin Ruhlin Gjuvsland, Senior Communication Consultant, Norwegian Computing Center

We decided to join BigInsight because it gave us an opportunity to take a deep dive into the data we already had, says Ellen Stavseth Paaske, Power Market Analyst in Hydro Energy.

Hydro and Norsk Regnesentral (NR) have collaborated on modelling spot prices on the power market for over 20 years. When the BigInsight opportunity appeared four years ago, Hydro saw that BigInsight could help make the price system even more advanced.

- We used to only work with historical weather, consumption and price data. BigInsight has given us the opportunity to look at other aspects of what determines the prices. For example, we now use both the bidding curve from the power exchanges and renewable energy as part of the model, says Ellen.

Hard targets

Alex Lenkoski, chief research scientist at NR, knows the energy market well. He has been working on modeling the electricity market since 2003.

- Even though Hydro and NR already had a good industrial connection, the BigInsight project is about developing further and fine-tuning the methods, Alex explains.

- In this project, we are working on a very complex and large data set, with advanced methods and with something that should yield a major industrial result. We worked hard for the first three years and a year ago the big turning point came: we found the key to the latest unknown factors.

Complex methods

Hydro is impressed by the complexity of the method. The method calculates a price for each hour every single day and uses a huge amount of data.

- We experience over and over again that our collaborators in BigInsight have the ability to come up with hypotheses, to challenge and then test them in an effective way, Ellen says.

Alex is glad Hydro recognizes the success of the BigInsight project.

- Here we prove that we are good at finding the most appropriate method for the data set we are working with, states Alex.

Competitive advantage for Hydro

Alex points out that the fruitful collaboration with Hydro is due to good communication throughout the progress of the project.

- The first thing we did was to find out if the data were usable to us. The fact that I have been able to verify the data with Hydro as the model has been created has enabled us to get the most out of the date set. Without this cooperation, the development would have gone much slower, says Alex.

For Hydro, having such an advanced model in house has given the company a competitive advantage, says Ellen.

- The fact that we are so close to this model means that we have greater control over the sensitivity of the market. We can also consider different scenarios to see how high or low the prices can go. Had we not been part of BigInsight, predicting the spot price would probably be much harder. Without BigInsight, we would have to rely on the historical curve, which gives a much less detailed view on how price is changing.

The constant development of these power markets makes it important to constantly update the model.

- We see that we get our data examined more thoroughly when we cooperate with BigInsight, explains Ellen. Alex and his team have knowledge of the latest models and new research in the field and this gives us a strong advantage.

We also know more about the outcome possibilities for the prices in the next two weeks. This allows an even better utilization of the market.

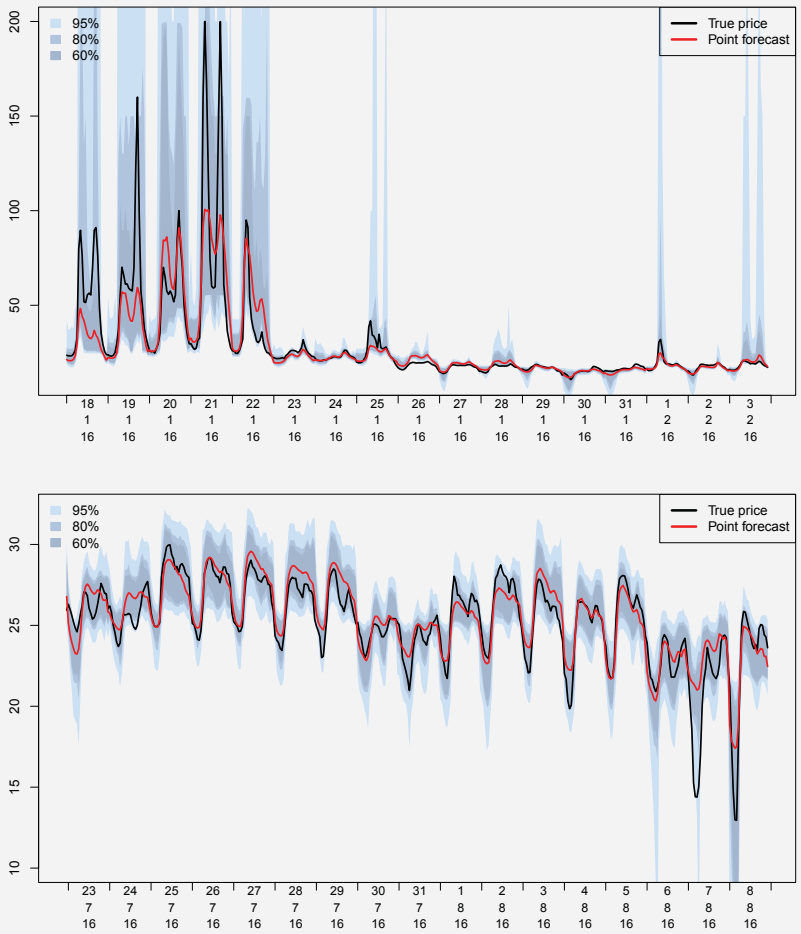


Figure: Probabilistic forecasts of electricity spot prices based on the bid/ask model for 17 days in January 2016 (upper row) and in July 2016 (lower row), displaying the 60%, 80%and 95% quantiles of the probabilistic distribution. The black line is electricity spot price pth and the red line is electricity spot forecasts p^th.

Extra gear

Success for researchers can be measured by a product getting a place in the market right away, Alex explains.

- Hydro is happy and we are happy. For me, working close to the industry with something so applied gives me an extra gear. Now that we've been able to solve the most difficult data set in this market, and can rely on what the model delivers, it's all about exploring the potential of our approach.

Hydro wants to improve the model also in other markets. Ellen believes Germany will represent an exciting opportunity:

- It is surprising how well it has actually worked on the Nordic market. The power market in Germany, where wind and solar power play an even greater role, will give this model a new challenge. We are excited to see if we can use this method in the German market, she concludes.

PERSONALISED MARKETING

We develop new methods, strategies and algorithms for individualised marketing, customer retention, optimised communication with users, personalised pricing and personalised recommendations or to maximise the probability of purchase of a product or other actions of the users. We exploit users' behavioural measurements in addition to their more standard characteristics and external data (including competitors' activity, market indicators, financial information, and geographic information). We exploit network topologies, informative missingness and temporal relations. A key point is to identify the actionable causes of customer behaviour.

What we did in 2018:

Stochastic customer growth dynamics

Understanding how networks of customers grow in time and topology is important. The Vipps transaction data may be viewed as a graph with users corresponding to the nodes and the financial transactions between the users defining the edges. With an advanced statistical model we have analysed the growth of this graph. Our experiments show that the intrinsic quality of the nodes plays an important part in the evolution of the network. This insight may be used to identify influential nodes for viral marketing. The approach has been accepted for publication in the journal Network Science.

Bankruptcy prediction using relational network data

Company defaults may be seen as a disease, spreading directly or indirectly from one company to others. Network relations are therefore assumed to play a fundamental role. We have first built a network between Norwegian enterprises where to companies are related if they share a director or a high-level manager. Then, we have computed a score for each company based on the bankruptcy behaviour of its neighbours in the network, and finally, we have added these network scores as additional variables in a traditional credit scoring model.

Bayesian methodology for recommender systems

Clicking data, which exist in abundance and contain objective user preference information, are widely used to produce personalized recommendations in web-based applications. Current popular recommendation algorithms, typically based on matrix factorizations, often achieve good click through rates. However, diversity of the recommended items, which can greatly enhance user experiences, is often overlooked. Moreover, most algorithms do not produce interpretable uncertainty quantifications of the recommendations. We propose the Bayesian Mallows for Clicking Data (BMCD), which leads to interpretable uncertainties of each individual recommendation, and we exploit these uncertainties in order to guide recommendations. We have worked on data from NRK TV. We demonstrate that compared to matrix factorization approaches, BMCD makes personalized recommendations with similar accuracy, while achieving a much higher level of diversity. In the next year, we will work on scalability of the algorithm and test it at finn.no. The methodology has been published on the Journal of Machine Learning Research and in several other journals.

Stochastic models for early prediction of viral customer behavior on networks

Can we predict if a new service or product will be go viral in a market, or if it will be a flop? We have developed a stochastic model that simulates the adoption of a product by users linked to each other by a social network, so that neighbours can convince each other (or discourage!). In 2018, we wrote very efficient algorithms for estimation and prediction and the method has been tested with success on Telenor's MinSky data. The next step is validating it further on new products and for different partners. We will also write a scientific paper, because the methodology is of generic value.

Clustering of clickstream data

Web stream data are routinely collected to study how users browse the web or use a service. The ability to identify user behaviour patterns from such data may be very valuable for different businesses. It may help to produce better marketing strategies, predict online purchases and a better user experience. We have used model-based clustering to segment users based on web clickstream data from Skatteetaten and Gjensidige. Model-based clustering assumes that users' behaviours are generated by a set of probabilistic models and each model corresponds to a cluster.

Principal Investigator
Kjersti Aas

co-Principal Investigator
Arnoldo Frigessi

PERSONALISED HEALTH AND PATIENT SAFETY

The health system is producing data at an unrestrainable speed; data that can mean personalized therapy, patient safety, personalized cancer prognoses, better prevention and monitoring of epidemics. We show how such data can be exploited, with a series of innovative projects.

What we did in 2018:

Personalized cancer statistics

National population based cancer registries publish survival statistics by cancer site, stage, gender and time period, using established epidemiological methods. As new clinical registries are established, more data on treatment and later events become available, in addition to information on comorbidity or income and educational level. Using cancer registry data, we are currently developing a set of models for testing when a population of cancer patients becomes "statistically cured", i.e. having the same risk of disease as the general population. This is important, both for testing the basic assumption of "statistical cure" made when extrapolating beyond the maximum follow-up times in the data, but also to be able to predict the length of the disease period for specific patient groups.

Estimating and predicting cancer drugs sensitivity in in-vitro screening

We have now a large project on drug sensitivity estimation and prediction, with several partners from the Oslo University Hospital. The aim is to be able to guide the selection of therapy based on the statistical prediction of how drugs will behave for the individual patient, each drug on its own or in combination. We have developed a new multivariate penalized regression method (IPF-tree-lasso) that improves prediction of drug sensitivity in large-scale screening experiments based on molecular characterization of cancer cell lines in two ways: by a more efficient combination of several sources of molecular data using the Integrative Lasso with Penalty Factors, which we extended to the IPF-elastic-net and (ii) by borrowing information from available data on similar drugs through a hierarchical tree structure in the penalty terms. In a second project we developed a new Bayesian approach to drug synergy estimation, based on thin plate splines. This allows understanding the effect of the doses of the two drugs on cell viability at a much finer scale. Inference is MCMC based. The next step will be to merge the two projects, work on Gaussian field approximations which allow scalability. We also started to develop a mathematical model of the drug screening experiment at cell level, which we hope can allow estimating the clonal composition of a cell sample.

Healthcare safety management

Can we estimate the risk of hospital acquired infections (HAI) in a hospital ward based on routinely measured clinical and organsational data? And if so, can we develop a tool that allows the hospital managers to understand the effect of possible interventions to reduce such risk. Electronic health records at patient level are time series or time processes of very varying type. Time series of the medical and nursing work in the ward are also available, together with external factors like air humidity and temperature. We have developed a new method, which is able to select which of these input time series affect most the output, namely a measure of the risk of HAI. The method has been tested on simulated data very succefully. We have not yet been allowed to work on the OUS data, and believe that a decision will be taken in 2019. We are approaching also other possible hospitals.

Network theory for health

The world is continuously urbanising, resulting in clusters of densely populated urban areas and more sparsely populated rural areas. We proposed a method for generating spatial fields with controllable levels of clustering of the population. We build a synthetic country, and use this method to generate versions of the country with different clustering levels. Combined with a metapopulation model for infectious disease spread, this allows us to in silico explore how urbanisation affects infectious disease spread. Under within-country restrictions on non-commuting travel, the final size decreases with increased population clustering. The effect of travel restrictions on reducing the final size is larger with higher clustering. The reduction is larger in the more rural areas. Within-country travel restrictions delay the epidemic, and the delay is largest for lower clustering levels. We implemented three different vaccination strategies vaccination (in space), preferentially vaccinating urban locations and preferentially vaccinating rural locations. The urban and uniform vaccination strategies were most effective in reducing the final size, while the rural vaccination strategy was clearly inferior. According to our model, within-country travel restrictions are likely to

be less effective in delaying epidemics, while they will be more effective in decreasing final sizes. In addition, to minimise final sizes, it is important not to neglect urban locations when distributing vaccines. To our knowledge, this is the first study to systematically investigate the effect of urbanisation on infectious disease spread and in particular, to examine effectiveness of prevention measures as a function of urbanisation.

Personalised cancer therapy guided by computer simulation

Mathematical modelling and simulation have emerged as a potentially powerful, time- and cost effective approach to personalised cancer treatment. In order to predict the effect of a therapeutic regimen for an individual patient, it is necessary to initialize and to parametrize the model so to mirror exactly this patient's tumor. We developed a comprehensive approach to model and simulate a breast tumor treated by two different chemotherapies in combination or not. In the multiscale model we represent individual tumor and normal cells, with their cell cycle and others intracellular processes (depending on key molecular characteristics), the formation of blood vessels and their disruption, extracellular processes, as the diffusion of oxygen, drugs and important molecules (including VEGF which modulates vascular dynamics). The model is informed by data estimated from routinely acquired measurements of the patient's tumor, including histopathology, imaging, and molecular profiling. We implemented a computer system which simulates a cross-section of the tumor under a 12 weeks therapy regimen. We showed how the model is able to reproduce patients from a clinical trial of OUS, both responders and not. We show that other drug regimens might have led to a different outcome. We were also able to correctly predict the outcome of one breast cancer patient.

Principal Investigator
Magne Thoresen

co-Principal Investigator
Clara Cecilie Günther

PERSONALISED FRAUD DETECTION

Fraud is expensive, affects common resources and prices and is therefore important to detect and prevent. Soft fraud, the exaggeration of legitimate insurance claims, is quite diffuse and difficult to spot. A sustainable welfare system and efficient insurance operations require the implementation of effective measures to limit fraud. Tax avoidance and tax evasion are other important types of fraud. We are also interested in money laundering detection. We develop adaptive tools that handle a diversity of data, including payment logs, relational networks, text and other available digital records, but under strict privacy protection regulations.

A further objective is to combine the multitude of fraud detection models in an optimal way, taking advantage of the strength of each predictor while blurring away weaknesses, and still obtaining coherent quantifications of the uncertainty in the fraud prediction. A similar objective is the development of new individualised anti-money laundering solutions. So far, the detection of suspicious transactions is based on labour-intensive semi-manual approaches and restricted to customers who significantly differ from the norm. Since the volume of banking transactions is steadily increasing, automated, intelligent tools are needed. The aim is to significantly increase the number of correctly identified money laundering transactions.

Fraud detection can be seen as a regression/forecasting problem, where fraud (true/false) is the response, possibly accompanied by a potential economic loss, and there are very many covariates. Including interactions, the number of covariates is huge. Generally, there are few fraud cases that are investigated, and a great number of undetected cases exist. The objective is to produce a trustworthy probability of fraud for each case.

What we did in 2018:

Combining dependent probability forecasts

Different fraud forecasting methods are likely to be dependent. This is in particular the case if they are based on the same covariates. Therefore, ensemble methods which ignore such dependence will treat very similar forecasts as independent. This will lead to an exaggerated confidence in the information from the individual forecasts. The idea is to construct a joint model for the outcome (fraud/not fraud) and the forecasts, based on a pair-copula construction, that captures the dependencies between these. We developed the pair copula methodology in the last decade, and it is now used in many areas with great success. The new method will be tested on simulated and real data from our partners.

Network analysis for fraud detection

Fraud can spread directly or indirectly from one fraudster to another. Exploiting knowledge about social relations between users/customers can be useful to discover fraud. Understanding how such networks of users look and evolve over time is expected to significantly improve fraud detection models. We build these networks and extract useful characteristics to produce better fraud forecasts and provide additional insight into how fraud spreads. We are here working with insurance fraud data, and later on tax avoidance and money laundering data.

A machine learning model for suspicious transactions

Most supervised anti-money laundering methods assume that suspicious activities are labelled as such by experts, while legitimate activities are just randomly sampled from the complete population of activities. This is motivated by the fact that the chance of a random activity being suspicious is almost zero. We challenge this view by 1) modelling suspicious transactions directly instead of via accounts or parties, and 2) show that the current practice of excluding activities labelled as non-suspicious by experts leads to significantly worse performance. The method is being transferred to DNB and a paper describing the approach has been submitted.

Local Gaussian discrimination with discrete and continuous variables

We generalise classical discriminant analysis (LDA and QDA) by replacing regular Gaussian distributions with local-Gaussian class distributions. This lifts the variable dependence from globally pairwise to locally pairwise. We are also able to combine discrete and categorical variables with continuous variables by relying on pairwise dependence in a unified framework. The method is evaluated on simulated data and real data from one of the partners, and a paper is submitted.

Principal Investigator
Anders Løland

co-Principal Investigator
Ingrid Hobæk Haff

SENSOR SYSTEMS

Sensor data are multidimensional streams of observations from any sensor system. In this IO we work mainly on sensor systems in the maritime sector, but as Statistics Norway has joined BigInsight in 2018, we consider their activity as ´sensoring´ society, and therefore include the research with SSB in this IO.

For maritime safety surveillance we develop new approaches based on the recent availability of large arrays of sensors, which monitor condition and performance of vessels, machinery and power systems. Sensor data are becoming increasingly available on global ship fleets, with efficient broadband connectivity to shore. We suggest new approaches to condition and/or performance monitoring, which is the process of identifying changes in sensor data that are indicative of a developing anomaly or fault. In addition to using previous failure data and pattern recognition techniques to detect anomalies, we test model based approaches that exploit knowledge on the sensors and the conditions they assess.

What we did in 2018:

Efficient on-line anomaly detection for ship systems in operation

For sensor based monitoring, methods for online (real time) fault and anomaly detection are important. Given the nature of multivariate sensor data streams from ships, considerable effort is put in neutralizing the context (operational mode etc.) in the data in real time, such that series of residuals, rather than the series themselves, are monitored. This step might be computationally heavy with long series of training data from normal conditions. We have improved considerably the computational time of a known method for residual generation (AAKR) introducing clustering techniques and other modifications. The improved on-line anomaly detection schemes have been tested on several multidimensional maritime sensor data sets from DNV-GL, and on bench mark data from other settings. This activity has resulted in several conference papers, and a full journal paper was published in 2018 in Expert Systems with Applications.

Anomaly detection for early warning in a motor cooling system

For a motor cooling system monitored by ABB, we have transferred a novel and highly efficient model based anomaly detection algorithm based on covariate information to the partner in the end of 2018. This detector gives a warning more than one hour before a known historical critical fault happens, and is now in production at ABB.

A virtual sensor system for hull condition monitoring

We have constructed a virtual sensor system (emulator) for the fatigue rate of a ship, given its physical characteristics and operating history, including environmental, sea state and weather data. The emulator is compared to calculations from a complex hydrodynamic model run by DNV-GL and will be used as a fast and precise supplement to the full model. Comparing results from the physical model and the emulator shows that the correspondence is adequate in most conditions. The system considerably simplifies data collection and calculations, and will be implemented at DNV-GL.

Also in connection to hull condition monitoring, we have collected real sensor data on bending moments from several vessels together with their history. The aim is to evaluate the precision of these measurements, comparing with corresponding results from the hydrodynamic model. First analyses comparing sensed bending and calculated bending for the same ships show that the correspondence is satisfactory in certain conditions, but not always.

Anomaly/fault classification and prediction from ship operation log files

One of the condition monitoring systems in a ship sends messages regarding the operational mode of the ship at irregular time points. This log file works on a finite alphabet of possible events, and the purpose is to analyse this log file to detect sequences of events which appear to be abnormal or correspond to failures, preferably giving a warning early enough to be able to take action. Based on a case from ABB, we have attacked this problem building on feature extraction, hidden Markov models and machine learning methods from speech recognition. Two papers are in preparation.

Sequential detection of changes using dimension reduction techniques for high dimensional sensor data

We have worked on novel methods for multi-stream sequential sparse change detection, in cases where the change appears simultaneously in a small, but unknown subset of sensors. A paper on how to tailor dimension reducing PCA for fast, online detection of various types of unknown sparse changes is in its final phase. Especially, the methodology is able to detect otherwise "invisible" changes in correlation between streams. The method is tested on comprehensive simulations, and will be applied

to DNV-GL and ABB ship data. Also, there is interest in trying out this change detector on data from other IO´s in BigInsight.

Realistic complexity in virtual test beds for autonomous vessel algorithms

We have initiated a new activity in the area of autonomous vessels, and in the end of 2018 we started to work on how to use Automatic Information System (AIS) data in order to create realistic test beds for autonomous vessel algorithms. This is important for the future classification of such vessels performed by DNV-GL. Combining expert knowledge and AIS data, we aim to describe virtual test beds where complexity and frequency of situations are realistic and at the same time span out the risks.

Dynamical linear models for condition monitoring with sensor data

As an alternative to AAKR for signal reconstruction under normal operations, we have used a number of multivariate dynamic linear models for the reconstruction, and combined with sequential testing of residuals. A journal paper was published in 2018 in International Journal of Condition Monitoring and Diagnostic Engineering Management.

Ship speed prediction based on full scale sensor measurements and environmental conditions

In this study we have evaluated various statistical models and machine learning algorithms to predict speed through water based on full scale sensor measurements, also of external forces. Such predictions can be used for performance monitoring and optimization of ship design and operation. A journal paper was published in 2018 in Ocean Engineering.

Data integration for improved commuting pattern estimation

With SSB we have started to explore how to use register data and Telenor's mobile phones' position data to estimate commuting patterns with a much finer spatial and temporal resolution, compared with today's estimates based on surveys. This work has been in the initial phase in 2018, with emphasis on data collection and organisation, and project formulations.

Principal Investigator
Ingrid Glad

co-Principal Investigator
Magne Aldrin

FORECASTING POWER SYSTEMS

Electricity producers rely on forecasts of electricity prices for bidding in the markets and power plant scheduling. Markets are changing: A much tighter integration between European markets and a rise in unregulated renewable energy production, especially wind and photo-voltaic, call for joint probabilistic forecasts. Incorporating the transient interplay between productions from renewable sources is critical to power production and financial operations. Multivariate probabilistic forecasts of electricity prices in the short horizon are required.

Appropriately characterising multivariate uncertainty will enable more effective operational decisions to be made.

Conventional power grids add extra generation and distribution capacity. Smart grids actively match energy supply and demand and combine the needs of the markets with the limitations of the grid infrastructure. With the implementation of smart meters and grid sensors, enormous amounts of time series data are generated, with seconds resolution. Our objective is to develop new methods that extract the right information from data to optimise grid control and for real time operation.

What we did in 2018:

Finalized Error dressing spot price forecast methodology

Electricity markets "spike" and "crash" when volumes are respectively slightly higher or lower than typical and these extreme price swings make uncertainty quantification a critical part of forecasts. However, the limited degree to which these extremes are observed makes such constructions difficult. In 2018 we finalized the system that uses published bid/ask curves that determine the final price to construct realistic distributional price forecasts that embed this extreme behaviour.

The model has been in the production environment at our industrial partner Hydro for over a year and the output is used in the decision making processes for their trading and production groups. In addition to continuing our collaboration with them, we have submitted our first major paper to the Annals of Applied Statistics discussing this methodology and results.

Finalized papers addressing the forecasting of renewable energy production

Having a good medium-term forecast of renewable energy production in Germany is becoming critical for actors in the power markets even in the Nordic countries. Fluctuations in German renewable production have a Europe-wide impact on prices. We finalized a long-running research project focused on Bayesian hierarchical regression that issues joint probabilistic forecasts of renewable energy production in Germany for one to seventy-two hours ahead. This has culminated in two papers, one addressing wind power production and another addressing sun (photo-voltaic) energy production. We submitted these two papers to IEEE Transactions on Power Systems.

Begun work on price spike warning systems

In building the Bid/Ask model mentioned above we interacted extensively with our industrial partner Hydro. The Bid/Ask model is quite general, in that it issues predictive distributions and it is up to the users to decide what to subsequently do with these distributions. The most pressing matter that Hydro used these distributions for was to construct a "warning system" that issues a probability that prices will exceed some high threshold.

The Bid/Ask methodology is only one manner that this warning system could be created. Given the importance and usefulness of the output to Hydro, we have begun a research project to directly focus on issuing probabilities of high-price spikes. In particular, we will compare the output of our bid-ask model to other more purpose-built classifiers. In 2018 we began the process of data collection for this research project.

Principal Investigator
Alex Lenkoski

co-Principal Investigator
Carlo Mannino

AI – EXPLANATION & LAW

At the intersection between artificial intelligence, transparency, privacy and law, there is a need for more research. This IO, which started up during 2018, consists of two parts: Explanation & Law.

Explanation: Artificial intelligence, statistical models or machine learning models can often be seen as black boxes to those who construct the model and/or to those who use or are exposed to the models. This can be due to: 1) Complicated models, such as deep neural nets, boosted tree models or ensemble models, 2) Models with many variables/parameters and 3) Dependencies between the variables.

Even simple models can be difficult to explain to persons who are not mathematically literate. Some models can be explained, but only through their global, not personalised, behaviour. There are a number of good reasons for explaining how a black box model works for each individual:

a. Those who construct or use the model should understand how the model works
b. Those who are exposed to the model should, and sometimes will, have the right to an explanation about a model's behavior
c. It should be possible to detect undesired effects in the model, for example an unfair or illegal treatment of certain groups of individuals, or too much weight on irrelevant variables.

Law: Research at BigInsight challenges some of the legal principles that govern data privacy, including the risk of re-identification of anonymised parties, the wish to minimise data made available to discover associations and causes and the uncertainty of the value created by big data research. The need for compromising between privacy protection and common good is particularly evident in medical research. Methods and algorithms should follow the five principles of responsibility, explainability, accuracy, auditability and fairness. How can these aspects be regulated, validated and audited? There is a need to design new legislation and legal practices that allow exploiting big data while guaranteeing privacy protection.

What we did in 2018:

Seminar series

We organized three seminars, where BigInsight researchers and partners presented and discussed their views on themes related to personalised explanations of personalised predictions: "Personalised explanations – what is it and why should we care about it?", "Personalised predictions: Ethics, anonymization, ownership of data and making data public" and "Methods for personalised explanations". Attendance and discussions were very god.

Correct explanations when there is dependence between the variables

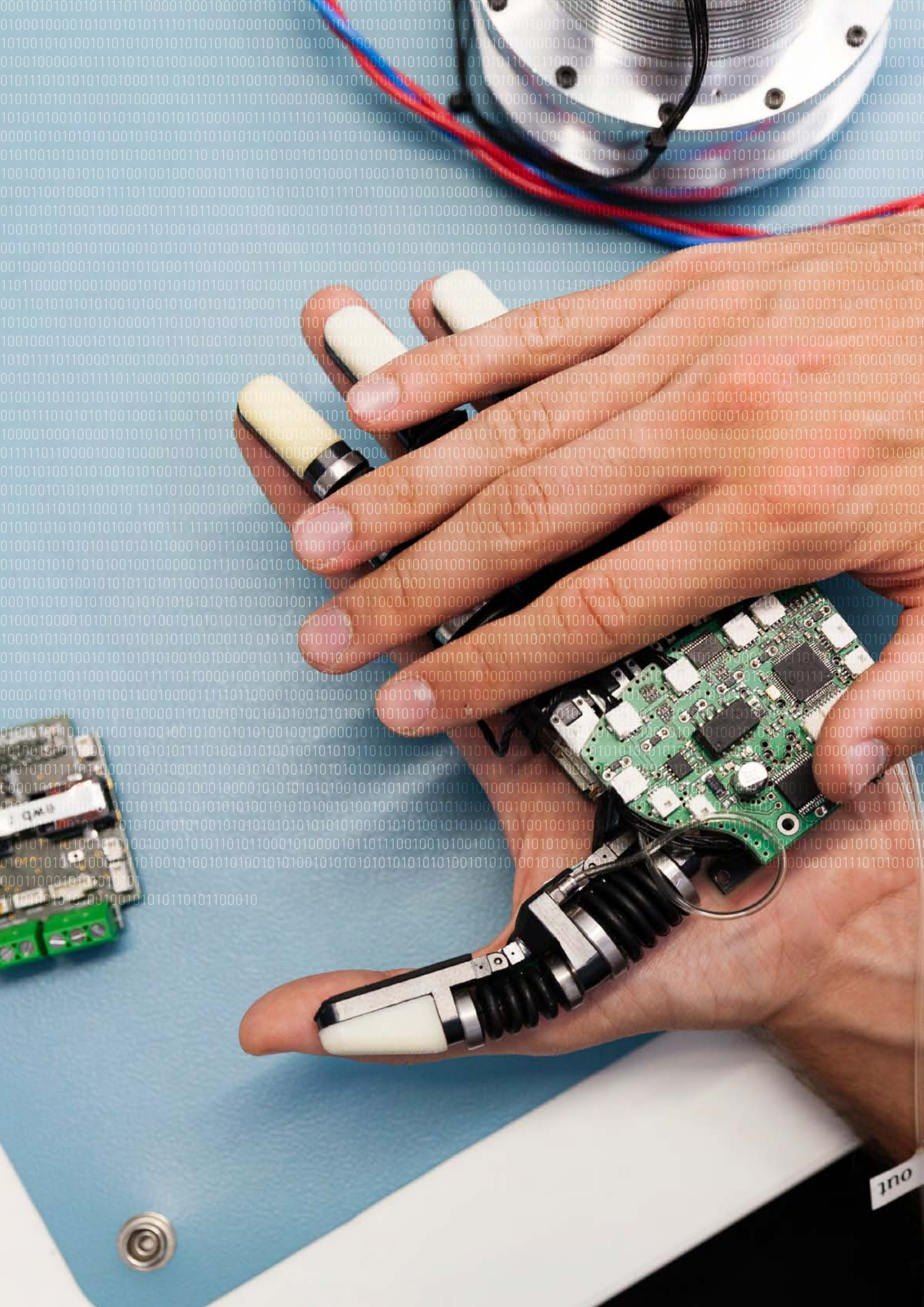
In many real life models, some or many of the variables of interest are dependent. For example, income and age typically follow each other quite closely. Current approaches to individual explanations do not handle dependent variables at all or not very well, especially in terms of the computational burden needed even for a handful of variables. We have been constructing new methods to handle these situations.

A PhD in law

We started to discuss with several groups of the Faculty of Law of the University of Oslo, the possibility to co-fund a PhD student on themes related to privacy, anonymization, fairness, and we believe that a call will be ready during 2019.

Principal Investigator
Anders Løland

co-Principal Investigator
Arnoldo Frigessi



INTERNATIONAL COOPERATION

International Academic Partners

International Academic Partners are key resources for BigInsight. We collaborate in research and co-supervise PhD students. We organize joint workshops and events.

STOR-i, Statistics and Operational Research in partnership with Industry, University of Lancaster

is a joint venture between the Departments of Mathematics & Statistics and Management Science of the University of Lancaster. STOR-i offers a unique interdisciplinary PhD programme developed and delivered with important UK industrial partners. The centre is at the forefront of international research effort in statistics and operation research, establishing an enviable track record of theoretical innovation arising from real world challenges. Professors Jonathan Tawn, professor Idris Eckley (who co-lead the centre) and professor David Leslie co-supervise PhD students together with BigInsight staff, on recommender systems, reinforced learning, multivariate extremes, non-parametric isotonic spatial regression, Bayesian modelling, multivariate sensor data, pair copula models. BigInsight and STOR-i co-organise industrial statistics sessions in international conferences and exchange membership in each other's scientific advisory boards. STOR-i has recently been renewed until 2023, also thanks to the strong links to BigInsight.

Professors Idris Eckley, Jonathan Tawn and Kevin Glazebrook, leading STOR-i at University of Lancaster

The Medical Research Council Biostatistics Unit (BSU)

is part of the University of Cambridge, School of Clinical Medicine. It is a major centre for research, training and knowledge transfer, with a mission 'to advance biomedical science and human health through the development, application and dissemination of statistical methods'. BSU's critical mass of methodological, applied and computational expertise provides a unique environment of cutting edge biostatistics, striking a balance between statistical innovation, dissemination of methodology and engagement with biomedical and public health priorities. Professor Sylvia Richardson is director of the BSU and she has received an honorary degree of the University of Oslo. BigInsight and the BSU have several joint projects in health and molecular biology. The prestigious Aker Scholarship has been awarded to master student Camilla Lingjærde who starts in 2019 a PhD at the BSU. We have also involved the BSU in our collaboration with the University of Hawassa (Ethiopia).

Professor Sylvia Richardson, MRC Biostatistics Unit, Cambridge

In 2018, BigInsight has also collaborated with **the Finnish Center for Artificial Intelligence, Aalto University** and the **Department of Statistics of the University of Chicago.**

International guest programme

BigInsight has an international guest programme, funding from short visits up to long term visiting and adjunct positions and a sabbatical visitor programme.

In 2018 we hosted the following longer visits:

Professor Gianpaolo Scalia Tomba, University of Roma Tor Vergata, visits Oslo regularly in collaboration with NIPH. He is interested in models for infectious diseases and antibiotics resistance.

Professor Emeritus Elja Arjas, University of Helsinki, has a 20% adjunct position at BigInsight and collaborates in projects and recommender systems and health.

Professor Jasmine Foo, University of Minnesota, has a sabbatical year (2018-2019) at BigInsight. She collaborates on BigInsight projects related to personalized cancer treatment, with particular focus on mathematical models for blood cancers. Fulbright Scholar Grant.

Ass. Professor Kevin Leder, University of Minnesota, has a sabbatical year (2018-2019) at BigInsight. He works with BigInsight on projects related to the mathematical study of drug synergies for cancer, including important optimization aspects. Fulbright Scholar Grant.

Other International activities

PhD students from other universities spent periods of training and research collaboration at BigInsight.

In 2018 we welcomed:

PhD Marta Crispino, University of Bocconi, Milano, has been successfully supervised to her PhD in Milano by Elja Arjas and Arnoldo Frigessi. Her thesis has been evaluated at highest level by an international committee. Marta is now postdoc in Grenoble, France, and continues her collaboration with BigInsight on recommender systems and rang data.

BigInsight is partner of the COST Action CA15109 "European Cooperation for Statistics of Network Data Science (COSTNET)".

Professor Arnoldo Frigessi is a member in the Management Committee and professor Birgitte Freiesleben de Blasio is nominated as deputy. This EU action started in 2016 and aims to facilitate interaction between diverse groups of statistical network modelers, establishing a large and vibrant interconnected and inclusive community of network scientists. The third workshop took place in Warsaw.

Dr. Geoff Canright, Telenor,
invited keynote speaker in Warsaw.

Scientific Advisory Committee of BigInsight

Scientific Advisory Committee of BigInsight has five international members. The next meeting will be in 2020.

Prof. Idris Eckley, Lancaster University, UK

- Until 2007 Statistical Consulant at Shell Global Solutions
- Co-Director of the EPSRC-funded STOR-i Centre for Doctoral Training
- Within STOR-i he leads the Centre's industrially-engaged research activity
- Co-Director of the Data Science Institute DSI@Lancaster: Lancaster's new world-class, multidisciplinary Data Science Institute.
- Leads the EPSRC programme StatScale: Statistical Scalability for Streaming Data

Prof. Samuel Kaski, University of Helsinki, Finland

- Professor of Computer Science, Aalto University
- Director, Finnish Centre of Excellence in Computational Inference Research COIN, Aalto University and University of Helsinki
- Academy Professor (research professor), 2016-2020
- Director, Finnish Center for Artificial Intelligence FCAI, 2018-
- Statistical machine learning and probabilistic modeling

Prof. Geoff Nicholls, University of Oxford, UK

- Professor in Statistics and Head of Department of Statistics
- PhD in particle physics in the Department of Applied Mathematics and Theoretical Physics in Cambridge, University of Auckland in New Zealand
- Bayesian inference, Computational Statistics, Statistic Genetics, Geoscience, Linguistics and Archaeology

Prof. Marina Vannucci, Rice University, Houston, USA

- Professor and Chair of the Department of Statistics
- Adjunct faculty member of the UT M.D. Anderson Cancer Center
- Rice Director of the Inter-institutional Graduate Program in Biostatistics
- Honorary appointment at the University of Liverpool, UK
- NSF CAREER award in 2001
- Former Editor-in-Chief for the journal Bayesian Analysis
- President, International Society for Bayesian Analysis

Reader Veronica Vinciotti, Brunel University of London, UK

- Ph.D in Statistics, Imperial College, London
- Research in statistical classification methods in credit scoring and in statistical genomics
- Co-director of the European Cooperation for Statistics of Network Data Science

PHD GRADUATES 2018

In 2018 the following PhD students affiliated to BigInsight defended their PhD thesis:

Marta Crispino
defended her PhD at Bocconi University, Milano, Italy:
Title of the thesis: Bayesian Learning of Ranking data.
Supervisors: Arnoldo Frigessi and Elja Arjas.
Co-advisor: Sonia Petrone (Bocconi)
External Reviewers: Jukka Corander (UiO) and David Leslie (U Lancaster)
Opponents: Carlo Baldassi and Daniele Durante (Bocconi)

Derbachew Asfaw
defended his PhD at University of Hawassa, Ethiopia.
Title of the thesis: Ranking Method and Inferences on the Admission Path of Students to University in Ethiopia.
Supervisors: Arnoldo Frigessi. Co-advisor: Zeytu Gashaw (U. Hawassa)
Opponents: Sebastian Engelke (University of Geneva, CH) and Stian Lydersen (NTNU)

ACTIVITIES AND EVENTS

2018 BigInsight Workshop

The yearly BigInsight Workshop was held on Monday 29th October at the headquarter of DNB. Representatives from the different partners, in addition to project members, were present. It was a successful event with a wide selection of presentations of our projects, ideas and results.

This year the Workshop started with breakfast and a morning dance. DJ Tveten was playing space disco, house and techno.

BigInsight Career Day

The first BigInsight career day took place on Wednesday 7th of November at Popsenteret. Most of the partners were present and gave a pitch of their company. More than 40 master, PhD and postdocs attended this successful event.

After the presentations, there was tapas and good drinks and all used the time to mingle and to make contacts.

Ingrid Glad was chairing the Career Day

Oslo Data Science Day

The University of Oslo Data Science Day on October 10th attracted more than 250 people and was a definite success with interesting presentations, stands, food, bar and mingling. The MatNat dean of innovation Kristin Vinje gave an opening speech and BigInsight PhD Sylvia Liu was chairing the whole event. Speakers were BigInsight´er assoc. prof. Valeria Vitelli and Facebook´s Valery Yundin.

PhD Sylvia Liu was chairing the event

TRAINING AND COURSES

BigInsight has been instrumental for the establishment of the new Master Program in Data Science at UiO, which started up in 2018. Admission to this master programme requires a bachelor with at least two statistics and two computer science courses, and as such it is different from many other competing programmes in Norway, which do not have such requirements. The focus of the master courses is on methods, algorithms, data analysis pipelines, in addition to project work and an internship in a data-intensive non-academic (private or public) environment. There is less focus on the use of available tools, because we believe that understanding the principles and foundations of data science is what will allow students to remain competent also in the future. In the first year, there were more than 250 applications and 15 students were admitted. BigInsight participates to the master by teaching, master projects and industrial contacts (for internship projects).

BigInsight staff supervise MSc projects in data science, and also in the more traditional master in statistics. When possible, we couple these projects to an on-going PhD project, so that the PhD student can participate to the supervision. PhD students also supervise MSc (and sometimes BSc) students who take a summer-job at NR or UiO.

Some PhD students work as teaching assistants, and in the final year also as teachers, in our courses, also at the Faculty of Medicine. Postdocs have teaching duties occasionally, and participate in supervision of master and bachelor students.

The Faculty of Mathematics and Natural Sciences soon starts an internship course for bachelor students across all fields of specialization. BigInsight staff has helped in the shaping of this course, using experience from contact with partners in the centre.

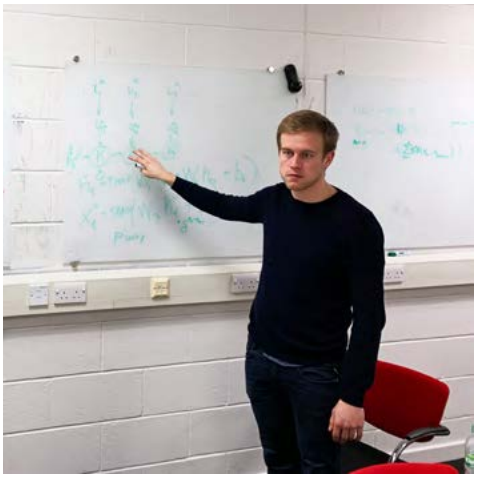
Thanks to BigInsight, there is a large cohort of PhD students at the Department of Mathematics, and at the Oslo Centre for Biostatistics and Epidemiology, which allows to organise more courses and activities for them. Several new PhD courses have been given (for example a course in methods for prediction, a second course in machine learning methods). Supervision of PhD students includes experts from the partners and the students often have direct and continuous contact with the partners.

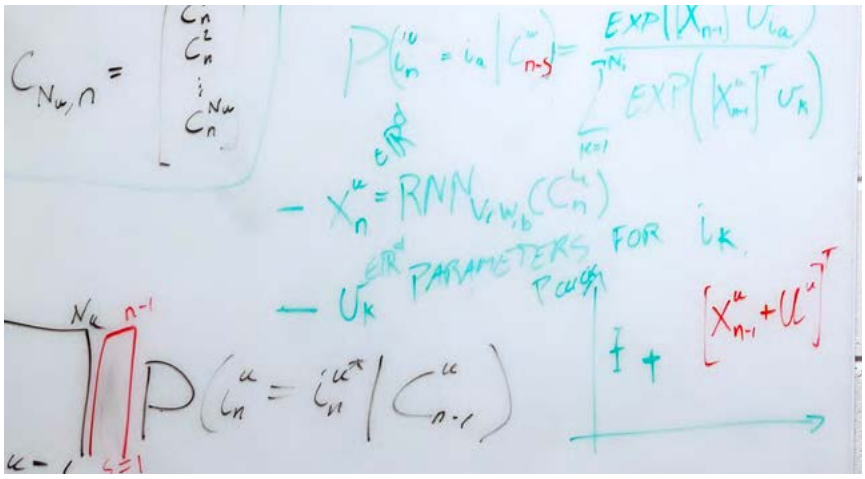
Of strong importance to the students society, BigInsight has fully or partly arranged and/or sponsored arrangements like Data Science Day, BigInsight Career Day, BigInsight Day and the yearly Klækken PhD workshop in 2018.

Many PhD students contribute to the advising services in statistics, biostatistics, bioinformatics and data science, which we offer to researchers at UiO and OUS. They follow an experienced advisor, before they advise on their own (with behind the scene support if needed). We offer a drop-in advising service and a more long term support. In this latter case, students are often co-authors of a research paper. These are very precious experiences. PhD students typically use about 2-3 weeks per semester in advising, on average.

Junior researchers at NR are mentored and participate in on-going BigInsight projects. This gives them an overview of the centre and a valuable exposure to methods and applications.


BigInsight
Oslo, Norway





COMMUNICATION AND DISSEMINATION ACTIVITIES

Website

The website of the center is biginsight.no.

Seminars

BigInsight co-organises the traditional Tuesday Statistical Seminar (at the Department of Mathematics) and the Thursday Biostatistics Seminar (at OCBE).

We also join forces with the Centre for Molecular Medicine Norway (NCMM) and launches the seminar series Sven Furberg Seminars in Bioinformatics and Statistical Genomics. The monthly seminars are joint bioinformatics-biostatistics catalyst events promoting scientific excellence and triggering collaborations on computational and statistical research projects related to molecular biology. The seminars are organized in three parts. First, a PhD student briefly presents their research. Second, the guest speaker gives a lecture on computational and/or statistical methods applied to molecular biology and medicine. Third, the audience gathers around pizza and refreshments. As part of the events, invited guest speakers meet local PIs and trainees.

BigInsight Wednesday Lunch Seminars are taken place every second week, alternating between the lunchroom at NR and the eight floor of the Department of Mathematics. While we share a good lunch, we listen to an invited lecture. Our speakers help us to understand global trends of data science developments of statistics, machine learning, operations research, optimisation, computer science, and mathematics in the era of high dimensional data.

BigInsight outreach presentations

EVENT / **ORGANISER**
Gjesteforelesning / **STK-MAT2011 UiO**
Credit Scoring using Deep Learning / **AIM2 North**
Seminar / **Women In Data Science**
Konferanse / **Make Data Smart Again 2018**
Konferanse / **DNB Data Summit 2018**
Workshop / **Forum for AI/data science i offentlig sektor**
Konferanse/Stage talk/discussion /
Technology Summit Kongsberg
Seminar GDPR for ledere /
Advokatfirmaet Schjødt og Infotjenester
Seminar / **NAV**
Frokostseminar / **DND**
Konferanse / **SESAR Innovation Days 2018, Salzburg**
Seminar / **Akers Solutions Innovation Week**
BigInsight for bigdata in eHealth /
UiO Faculty of Medicine board
Workshop on BigInsight`s work on sensors / **DNV-GL**
Fra data til innsikt Innovasjonskonferanse 2019 / **UiO**
Cutting Edge Festival 2018 /
Oslotech, Inven2, Conventor and UiO
Lederseminaret OSS 2018 / **OUS**
Arkitekter i DNB workshop / **DNB, Popsenteret**

BigInsight in the media

Bioteknologiradet.no, 14.02.2019, "Temaside til ungdomskoleprosjekt: 'Gen'-bokas hemmeligheter», Arnoldo Frigessi
Forskning.no, 14.02.2019, «Enkelte forskere går seg vill i frykten for datatørke», Anders Løland
Podcast LØRN.TECH #200, 18.01.2019, «AI – Keiserens kunstige klær», André Teigland
Apollon, 24.01.2019 "Simulerer kreftbehandlingen din på datamaskinen", Yngve Vogt og Arnoldo Frigessi
Apollon.uio.no, 14.02.2019 «Simulating your cancer treatment on a computer», Yngve Vogt og Arnoldo Frigessi
Apollon, 24.01.2019 «Unngår skipskatastrofer med avansert statistikk», Yngve Vogt og Ingrid Glad
Apollon.uio.no, 06.02.2019 «Avoids ship disasters with advanced statistikk», Yngve Vogt and Ingrid Glad
Apollon, 14.02.2019 «Usikker fremtid for svindlere», Yngve Vogt og Ingrid Hobæk Haff
Apollon.uio.no, 14.02.2019 «Uncertain future for fraudsters» Ingrid Hobæk Haff
ScienceNordic.com, 14.02.2019 "Uncertain future for fraudsters", Yngve Vogt and Ingrid Hobæk Haff
ScienceNordic.com, 14.02.2019 "Simulating your cancer treatment on a computer", Yngve Vogt and Arnoldo Frigessi
Dagens Næringsliv, 12.01.2018 «Kunstig intelligens til å stole på?» Anders Løland, Kari Laumann, Bjørn Erik Thon
Dagens Næringsliv, 03.02.2018 «Robot vs. folk flest: like mange bom» Anders Løland
NRK Ytring, 20.02.2018 «Ikke se til Facebook!» Anders Løland
Dagens Næringsliv, 08.06.2018 «Virkelige vansker og roboter» Anders Løland
Aftenposten, 27.06.2018 «Norsk næringsliv trenger flere IKT-eksperter» Randeberg, Røttingen og Baastad (BigInsight mentioned in the article)
Aftenposten, 19.08.2018 «Fem utfordringer for ansvarlig digitalisering» Anders Løland
Dagens Næringsliv, 07.09.2018 «Algoritmer er ikke alltid rettferdige» Anders Løland

Titan.uio.no, 24.09.2018 «Bruker datasimulering for å finne den beste kreftbehandlingen» interview with Arnoldo Frigessi
Centre for Digital Life Norway, 04.10 2018 «- The digital doctor will see you now» Can a computer predict disease and give advice on which treatment is best for you?
forskning.no, 16.10.2018 «For å forske med kunstig intelligens og maskinlæring må metodene tilpasses virkeligheten» Løland, Anders; Jarbekk, Eva; Skovly, Christian Werner
Podcast LØRN.TECH #48, 23.10.2018 «Ansvarlig bruk av stordata - Denne låneroboten kan avgjøre om du er en trygg eller risikabel kunde» Anders Løland
UniOslo, 01.11.2018 «PERCATHE has a new approach to cancer therapy». Arnoldo Frigessi and the convergence environment PERCATHE are developing a new approach to personalized cancer therapy, based on mathematical, statistical and biological methods.
Dagens næringsliv, 30.11.2018 «En sak for algoritmetilsynet» Anders Løland

RECRUITMENT

BigInsight's partners recruit researchers, postdocs, PhD students, Master students and summerstudents, in order to staff our projects. This happens with funding both from BigInsight and associated projects.

Those who started in 2018 were:

NAME	POSITION	FUNDING	RESEARCH AREA	AFFILIATION
Azzeddine Bakdi	Postdoc	BigInsight	Sensor	UiO
Simon Boge Brant	PhD student	BigInsight	Fraud	UiO
Simen Eide	PhD student	Finn.no	Marketing	UiO
Jaroslaw Nowak	PhD student	ABB	Sensor	ABB
Riccardo Parviero	PhD student	BigInsight	Marketing	UiO
Brittany Rose	PhD student	BigInsight/NIPH	Health	NIPH
Leiv Tore Salte Rønneberg	PhD student	BigInsight	Health	UiO
Jonas Fredrik Schenkel	PhD student	BigInsight	Sensor	UiO

PERSONNEL

Personnel affiliated with BigInsight for at least 10% of their time.

NAME	INSTITUTION	MAIN RESEARCH AREA
Arnoldo Frigessi	UiO/OUS/NR	Marketing, Health, Sensor, AI Explanation & Law
Jaroslaw Novak	ABB	Sensor
Morten Stakkeland	ABB	Sensor
Bjørn Møller	OUS/CRN	Health
Jan Nygård	OUS/CRN	Health
Johannes Lorentzen	DNB	Fraud
Fredrik Johannessen	DNB	Fraud
Karl Aksel Festø	DNB	Marketing
Marcus Nilsson	DNB	Marketing
Aiko Yamashita	DNB	AI Explanation & Law
Nafiseh Shabib	DNB	Marketing
Hodjat Rahmati	DNB	Marketing
Roger Olafsen	DNB	Marketing
Geir Ånonsen	DNB	Fraud
Ole Christian Astrup	DNV-GL	Sensor
Håvard Nordtveit Austefjord	DNV-GL	Sensor
Andreas Brandsæter	DNV-GL	Sensor
Ørjan Fredriksen	DNV-GL	Sensor
Gaute Storhaug	DNV-GL	Sensor
Bjørn-Johan Vartdal	DNV-GL	Sensor
Erik Vanem	DNV-GL	Sensor, Poser
Anders Nyberg	Gjensidige	Marketing
Geir Thomassen	Gjensidige	Fraud
Knut-Harald Bakke	Hydro	Power
Ellen Paaske	Hydro	Power

NAME	INSTITUTION	MAIN RESEARCH AREA
Peter Szederjesi	Hydro	Power
Birgitte De Blasio	NIPH	Health
Christopher S. Nielsen	NIPH	Health
Ulf Andersen	NAV	Fraud
Robindra Prabhu	NAV	AI Explanation & Law
Kjersti Aas	NR	Marketing
Magne Aldrin	NR	Sensor
Clara-Cecilie Günther	NR	Marketing, Health
Ola Haug	NR	Marketing, Sensor
Marion Haugen	NR	Marketing
Kristoffer Herland Hellton	NR	Marketing, Sensor
Lars Holden	NR	Health, Fraud
Ragnar Bang Huseby	NR	Fraud, Power
Martin Jullum	NR	Fraud
Alex Lenkoski	NR	Power
Anders Løland	NR	Fraud, Power
Linda R. Neef	NR	Fraud
Hanne Rognebakke	NR	Marketing, Sensor
Nikolai Sellereite	NR	Marketing
Gunnhildur Steinbakk	NR	Fraud, Sensor, Power
André Teigland	NR	Marketing, Fraud, Power
Ingunn Fride Tvete	NR	Health
Annabelle Redelmeier	NR	AI Explanation & Law
Mette Langaas	NR/NTNU	Sensor, Health
Torsten Eken	OUS	Health
Eivind Hovig	OUS	Health
Vessela Kristensen	OUS	Health
Marissa LeBlanc	OUS	Health
Sygve Nakken	OUS	Health
Andrew Reiner	OUS	Marketing, Health
David Swanson	OUS	Health
Irena Jakopanec	OUS	Health
Anders Berset	Skatteetaten	Marketing, Fraud
Wenche Celiussen	Skatteetaten	Marketing, Fraud
Øystein Olsen	Skatteetaten	Marketing
Nils Gaute Voll	Skatteetaten	Marketing, Fraud
Anders Holmberg	SSB	Marketing, Sensor, Power
Øyvind Langsrud	SSB	Marketing, Sensor, Power
Kim Benjamin Boué	SSB	Marketing
Li-Chun Zhang	SSB	Sensor
Kenth Engo-Monsen	Telenor	Marketing, Health, Sensor
Geoffrey Canright	Telenor	Marketing, Health AI E&L
Bård Støve	UiB	Fraud
Dag Tjøstheim	UiB/NR	Fraud
Elja Arjas	UiO	Marketing
Ørnulf Borgan	UiO	Marketing
Jukka Corander	UiO	Health
Ingrid K. Glad	UiO	Sensor
Ingrid Hobæk Haff	UiO	Fraud
Nils Lid Hjort	UiO	Fraud, Sensor
Carlo Mannino	UiO	Power
Kjetil Røysland	UiO	Health
Ida Scheel	UiO	Marketing
Geir Storvik	UiO	Sensor
Magne Thoresen	UiO	Health
Manuela Zucknick	UiO	Health

NAME	FUNDING	NATIONALITY	PERIOD	GENDER	TOPIC
Postdoctoral researchers with financial support from BigInsight					
Marta Crispino		Italy	2017-2018	F	Marketing
Alvaro Köhn Luque		Spain	2016-2020	M	Health
Azzeddine Bakdi		Algeria	2018-2021	M	Sensor
Postdoctoral researchers in BigInsight with financial support from other sources					
Andrea Cremaschi	UiO/NCMM	Italy	2016-2018	M	Health
Gudmund Hermansen	UiO	Norway	2016-2018	M	Sensor
Christian Page	OUS/HSØ	Norway	2016-2018	M	Health
Øystein Sørensen	UiO	Norway	2018	M	Marketing
Owen Thomas	UiO	UK	2017-2019	M	Health
Valeria Vitelli	UiO	Italy	2017-2020	F	Health, Marketing
PhD students with financial support from BigInsight					
Simon Boge Brant		Norway	2018-2021	M	
Solveig Engebretsen		Norway	2016-2019	F	Health
Emanuele Gramuglia		Italy	2016-2019	M	Sensor
Andrea Chi Zhang		China	2016-2019	F	Health
Christoffer Haug Laache		Norway	2017-2020	M	Health
Brittany Rose		USA	2018-2021	F	Health
Leiv Tore Salte Rønneberg		Norway	2018-2021	M	Health
Riccardo Parviero		Italy	2018-2021	M	Marketing
Jonas Schenkel		Norway	2018-2021	M	SSB, Sensor
Martin Tveten		Norway	2017-2020	M	Sensor
PhD students in BigInsight with financial support from other sources					
Derbachew Asfaw	UiO	Ethiopian	2016-2018	M	Marketing
Andreas Brandsæter	DNV-GL, NæringslivPHD	Norway	2015-2018	M	Sensor
Simen Eide	Finn.no, NæringslivPhD	Norway	2018-2021	M	Marketing
Vinnie Ko	UiO	Netherlands	2016-2019	M	Fraud
Håvard Kvamme	UiO	Norway	2015-2018	M	Marketing
Jaroslaw Nowak	ABB	Poland	2018-2021	M	Sensor
Richard Xiaoran Lai	UiO	UK	2015-2018	M	Health
Sylvia Qinghua Liu	UiO/MI Innovation	China	2016-2020	F	Marketing
Andreas Nakkerud	UiO/MI Innovation	Norway	2016-2020	M	Power
Zhi Zhao	UiO/IMB	China	2016-2018	M	Health
Yinzhi Wang	UiO/MI	China	2016-2018	F	Fraud
Master degrees					
Kristin Bakka	NTNU	Norway	2017-2018	F	Sensor
Jenine Gaspar Corrales	NR	Spain	2017-2018	F	Marketing

FINANCIAL OVERVIEW

FUNDING	1000 NOK
The Research Council	12 587
Norwegian Computing Center (NR)	1 988
Research Partners*, in kind	10 508
Research Partners*, in cash	762
Enterprise partners**, in kind	3 858
Enterprise partners**, in cash	4 341
Public partners***, in kind	4 825
Public partners***, in cash	1 957
Sum	**40 826**

COSTS	
NR, research	10 216
NR, direct costs	718
Research Partners*, research	20 811
Enterprise partners**, research	3 858
Public partners***, research	5 225
Sum	**40 826**

*Research partners: UiO, UiB
** Enterprise partners: Telenor, DnB, Gjensidige, Norsk Hydro, DNV-GL, ABB
*** Public partners: Norwegian Tax Administration (Oslo), University Hospital HF, NAV, Public Health Institute (NIPH), Statistics Norway

PUBLICATIONS IN 2018 incl up to feb 2019

Journal and peer-reviewed conference papers

Aas K, Rognebakke H. The evolution a mobile payment solution network, to appear in Network Science 2018.

Brandsæter A, Vanem E, Glad IK. (2019) Efficient on-line anomaly detection for ship systems in operation. Expert Systems with Applications, Vol.121, pp. 418-437. Available online December 22, 2018.

Brandsæter A, Vanem E. Ship speed prediction based on full scale sensor measurements of shaft thrust and environmental conditions. Ocean Engineering, 2018, 162, 316-330.

Crispino M, Arjas E, Vitelli V, Barrett N, Frigessi A. A Bayesian Mallows approach to non-transitive pair comparison data: How human are sounds? To appear Ann Appl Statistics, 2018.
Editor pick: Mogil JS. Friends in pain: pain tolerance in a social network. Scand J of Pain. 2018.

Engebretsen S, Glad IK. (2019) Additive Monotone Regression in Lower and Higher Dimensions. Accepted for publication in Statistics Surveys.

Engebretsen S, Frigessi A, Engø-Monsen K, Furberg AS, Stubhaug A, de Blasio BF, Nielsen CS. The peer effect on pain tolerance. Scandinavian Journal of Pain. 2018.

Engebretsen S, Engø-Monsen K, Frigessi A, Freiesleben de Blasio B (2019) A theoretical single-parameter model for urbanisation to study infectious disease spread and interventions. PLoS Comput Biol 15(3): e1006879. https://doi.org/ 10.1371/journal.pcbi.1006879

Fanaee-T H, Thoresen M. Multi-insight visualization of multi-omics data via ensemble dimension reduction and tensor factorization. To appear Bioinformatics (2018).

Ghosh A, Thoresen M. Non-Concave Penalization in Linear Mixed-Effects Models and Regularized Selection of Fixed Effects. Advances in Statistical Analysis 2018, 102: 179-210.

Grytten I, Rand KD, Nederbragt AJ, Storvik GO, Glad IK, Sandve GK. Graph Peak Caller: calling ChIP-Seq Peaks on Graph-based Reference Genomes, Accepted PLOS COMP Biology, 2018.

Kvamme H, Sellereite N, Aas K, Sjursen SAS. Predicting mortgage default using convolutional neural networks. Expert systems with applications. 102 (2018): 207-217.

LeBlanc M, Zuber V, Thompson WK, Andreassen OA, Frigessi A, Andreassen BK. A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework. BMC genomics. 2018 Dec;19(1):494.

Lien TG, Borgan Ø, Reppe S, Gautvik K, Glad IK. Integrated analysis of DNA-methylation and gene expression using high-dimensional penalized regression: a cohort study on bone mineral density in postmenopausal women, BMC medical genomics. 2018 11 (1), 24.

Liu Q, Crispino M, Scheel I, Vitelli V, Frigessi A. Model-Based Learning from Preference Data. Annual Review of Statistics and Its Application. 2018.

Rand KD, Grytten I, Nederbragt AJ, Storvik GO, Glad IK, Sandve GK. Coordinates and intervals in graph-based reference genomes, BMC Bioinformatics. 2018 18 (1), 263.

Rohrbeck C, Costain DA, Frigessi A. Bayesian spatial monotonic multiple regression. Biometrika, 2018, 105.3: 691-707.

Rohrbeck C, Eastoe EF, Frigessi A, Tawn JA. Extreme value modelling of water-related insurance claims, Annals of Applied Statistics 2018, Vol. 12, No. 1, 246-282.

Sørensen Ø, Herland Hellton K, Frigessi A, Thoresen M. Covariate selection in high-dimensional generalized linear models with measurement error, To appear J. Computational and Graphical Statistics (2018)
With R-package hdme: High-Dimensional Regression with Measurement Error https://cran.r-project.org/web/packages/hdme/index.html.

Vanem E. Statistical methods for condition monitoring systems. International J. Condition Monitoring. 2018. 8(1) pp 9-23.

Vitelli V, Sørensen Ø, Crispino M, Frigessi A, Arjas E. Probabilistic preference learning with the Mallows rank model, Journal of Machine Learning Research 18 (2018) 1-49.
With R-package: "BayesMallows", https://cran.r-project.org/package=BayesMallows.

Reports and submitted papers

Cremaschi, A, Frigessi A, Tasken K, Zucknick M. A Bayesian approach for the study of synergistic interaction effects in drug combination in-vitro experiments, submitted 2019.

Divino F, Belay DB, Keilman N, Frigessi A. Bayesian Modelling of Lexis Mortality Data, arXiv 2018, submitted European J of Population.

Fanaee-T H, Thoresen M, Dimension reduction methods for integrative analysis of mRNA expression and DNA methylation: A benchmarking study. Submitted.

Gramuglia E, Storvik G, Stakkeland M. Fault prediction and classification for categorical streaming data. Submitted to JRSS Series C.

Jullum M, Løland A, Huseby RB, Ånonsen G, Lorentzen JP. Detecting money laundering transactions – which transactions should we learn from? Submitted to Expert Systems with Applications, 2018.

Kvamme H, Scheel I, Borgan Ø. Time-to-Event Prediction with Neural Networks and CoxRegression, submitted 2018 J Machine Learning Research.

Lai X, Geier O, Fleischer T, Garred O, Borgen EF, Funke S, Kumar S, Rognes ME, Seierstad T, Boerresen-Dale A-L, Kristensen VN, Engebraaten O, Kohn-Luque A, Frigessi A. Towards personalized computer simulation of breast cancer treatment: a multi-scale pharmacokinetic and pharmacodynamic model informed by multi-type patient data. bioRxiv 2018, in revision for Cancer Research.

Menden MP, Wang D, Guan Y, Mason M, Szalai B, Bulusu KC, Yu T, Kang J, Jeon M, Wolfinger R, Nguyen T, Zaslavskiy M, AstraZeneca-Sanger Drug Combination DREAM Consorti, Jang IS, Ghazoui Z, Ahsen ME, Vogel R, Neto EC, Norman T, Tang EKY, Garnett MJ, Di Veroli D, Fawell S, Stolovitzky G, Guinney J, Dry JR, Saez-Rodriguez J. A cancer pharmacogenomic screen powering crowd-sourced advancement of drug combination prediction, 2018, bioRxiv 200451.

Otneim H, Jullum M, Tjøstheim D. Local Fisher and pairwise naive Bayes: Improving two standard discriminants. Journal of Econometrics, special issue on Statistical Learning of Dependent Data, Submitted 2018 (invited paper).

Romeo G, Thoresen M. Model selection in high-dimensional noisy data: a simulation study. In revision for Statistical computation and simulation.

Salvatore S, Rand KD, Grytten I, Ferkingstad E, Domanska D, Holden L, Gheorghe M, Mathelier A, Glad IK, Sandve GK. (2018) Robustness of metrics for assessing similarity of genomic datasets. Submitted to Bioinformatics.

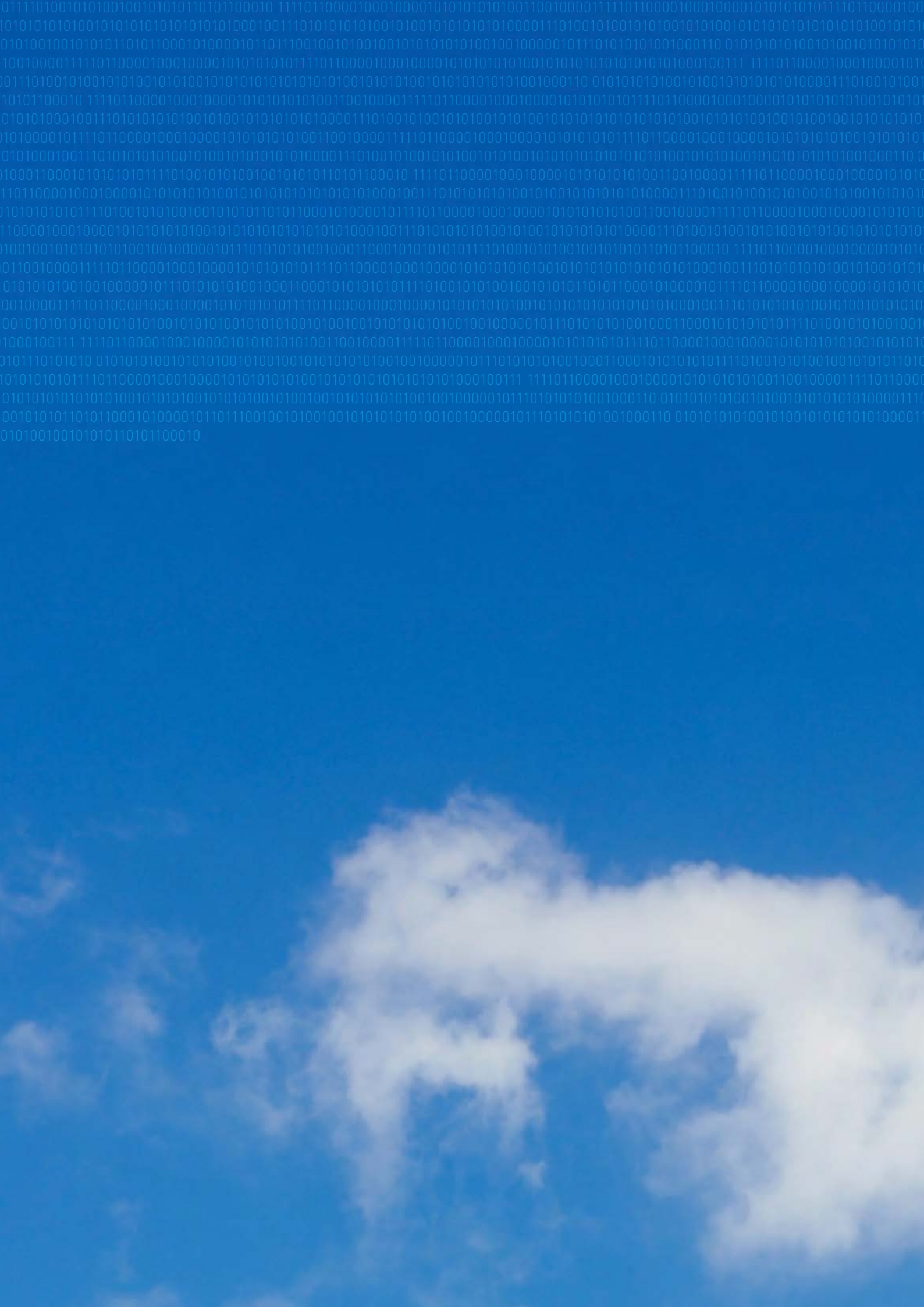
Steinbakk HG, Lenkoski A, Huseby RB, Løland A, Øigard TA. Using Published Bid/Ask Curves to Error Dress Spot Electricity Price Forecasts, submitted Annals Applied Statistics, 2018.

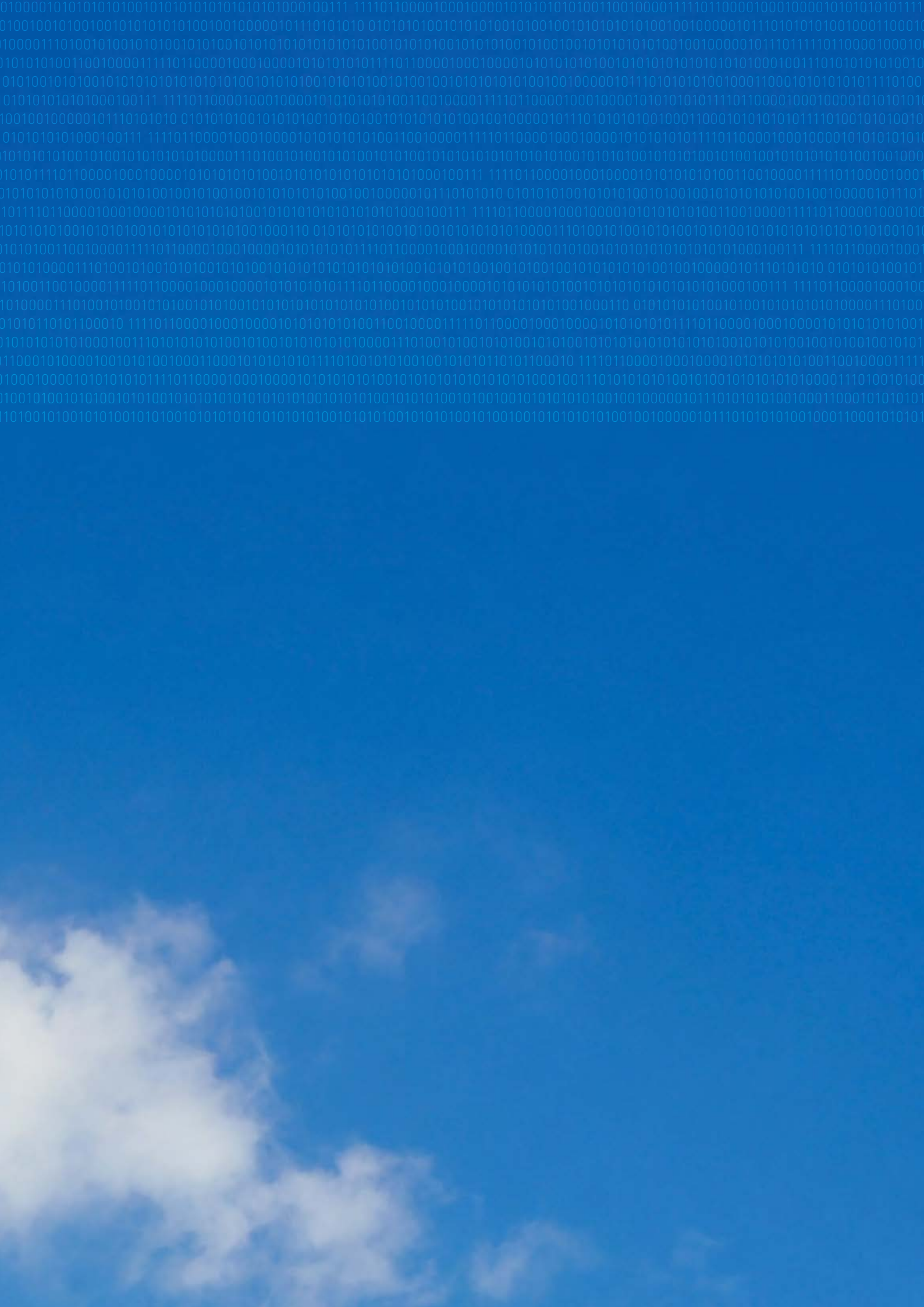
Swanson D, Lien T, Bergholtz H, Sorlie T, Frigessi A. (2018). A Bayesian Two-Way Latent Structure Model for Genomic Data Integration Reveals Few Pan-Genomic Cluster Subtypes in a Breast Cancer Cohort, bioRxiv 2018, in revisions for Bioinformatics.

Thorarinsdottir, T, Løland, A, Lenkoski, A. Probabilistic Forecasting of Temporal Trajectories of Regional Power Production – Part 1: Wind, submitted to IEEE Transactions on Power Systems, 2018.

Thorarinsdottir, T, Løland, A, Lenkoski, A. Probabilistic Forecasting of Temporal Trajectories of Regional Power Production – Part 2: Photovoltaic Solar, submitted to IEEE Transactions on Power Systems, 2018.

Zhao Z, Zucknick M. Incorporating correlations between drugs and heterogeneity of multi-omics data in structured penalized regression for drug sensitivity prediction. Submitted to Biostatistics, 2019.





BigInsight

Biginsight.no

Postal address
PO box 114 Blinderen
NO-0314 Oslo
Norway

Visiting addresses
BigInsight
Norsk Regnesentral
Gaustadalléen 23a
Kristen Nygaards hus, 4th floor
0373 Oslo

BigInsight
Oslo Center for Biostatistics and Epidemiology (OCBE)
University of Oslo
Sognsvannsveien 9
Domus Medica
0372 Oslo

BigInsight
Department of Mathematics
University of Oslo
Moltke Moes vei 35
Niels Abel Hus, 8th floor
0316 Oslo

BigInsight
Oslo Center for Biostatistics and Epidemiology (OCBE)
Oslo University Hospital
Klaus Torgårdsvei 3
Sogn Arena, 2nd floor
0372 Oslo

Email contacts
Arnoldo Frigessi frigessi@medisin.uio.no
Ingrid Glad glad@math.uio.no
Lars Holden lars.holden@nr.no
Ingrid Hobæk Haff ingrihaf@math.uio.no
André Teigland andre.teigland@nr.no
Kjersti Aas kjersti.aas@nr.no
Anders Løland anders.loland@nr.no

Phone contact
Arnoldo Frigessi +47 95735574
Norsk Regnesentral +47 22852500