BigInsight

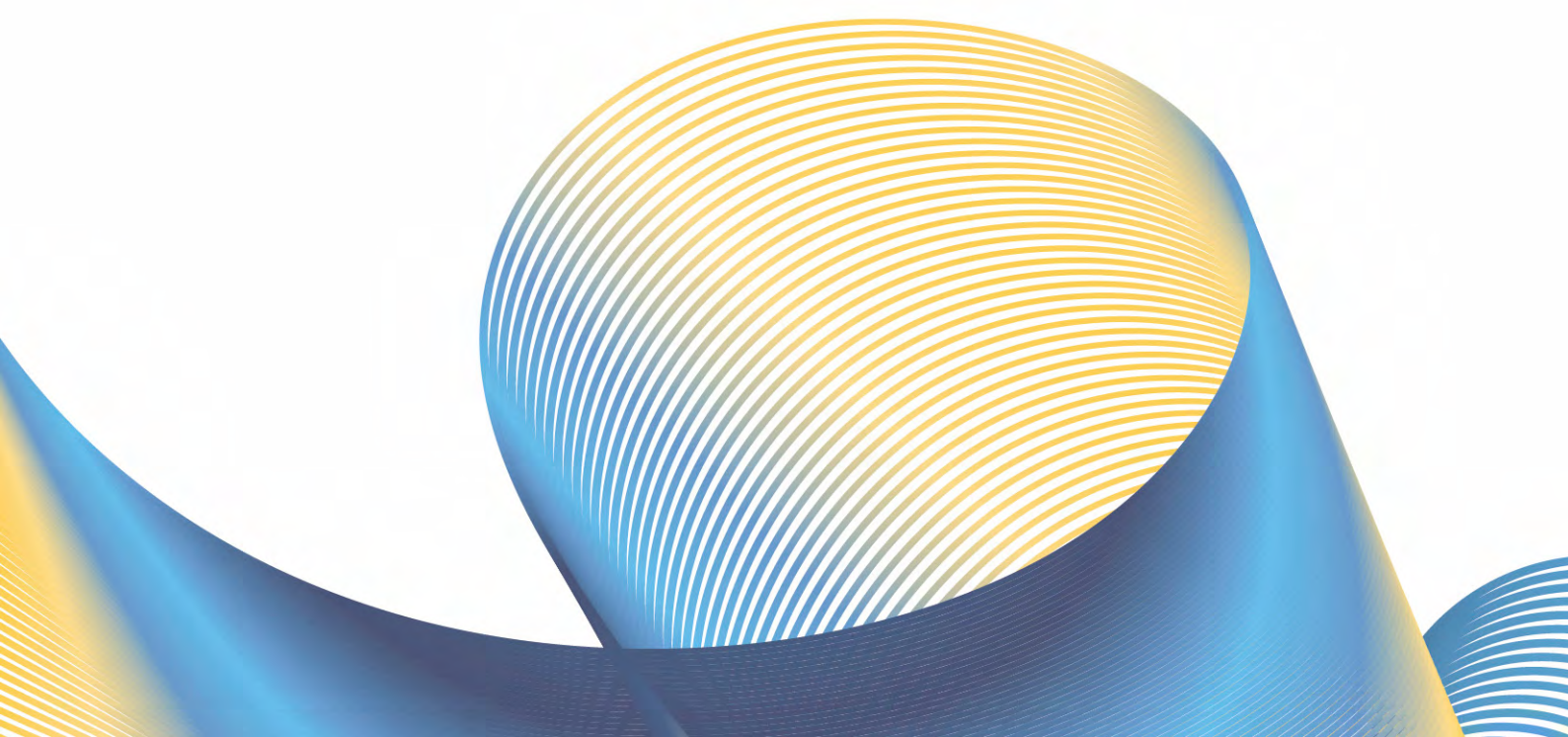
STATISTICS FOR THE KNOWLEDGE ECONOMY

ANNUAL REPORT 2022

Norwegian Centre
for Research-based
Innovation

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 problems facing our partners, by developing original statistical and machine learning methodologies.

Exploiting unique data resources and substantial scientific and domain specific knowledge, we construct personalised solutions, predict dynamic behaviours and control processes that are at the core of the partners' innovation strategies, and more generally of contemporary AI. 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 special situations. 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.

Methods and algorithms we develop and implement at BigInsight are explainable, accurate and fair. Our research is open. Research at BigInsight leads to value creation and strengthens 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 seventh year of BigInsight. Innovation results are highlighted, together with the broad spectrum of research projects.

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. The value does not reside in the data alone, but in the methods to extract knowledge from them.

Digitalisation means producing data, organizing and storing data, accessing data and analyzing data. BigInsight works in this last direction. 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 statistics and machine learning. We develop original, cutting-edge statistical, mathematical and machine learning methods, produce high-quality algorithms implementing these approaches and thereby deliver new, powerful, operational solutions. 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 courageous and creative, exploit knowledge and 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.

Personalised solutions

The core operation of our partners involves interacting with many individual units: customers, users, patients, but also sensors, vessels, wind-turbines, etc. Beside their obvious differences, there are many common characteristics:

- the high number of units/individuals/sensors under consideration;

"Kunstig intelligens som er klar om tre–fire år vil hjelpe oss å gjøre helse og utdanning mye bedre."

sier Bill Gates til VG, 14.2.23

- in some cases, massive data for each unit; in other cases, more limited information per unit;
- complex dependence structure between units;
- new data types, new technologies, new regulations make their use innovative;
- in most cases, units have their own intelligence, their own strategies and are exposed to their specific 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, personalised 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

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. Our partners 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 health-care institutions allows completely new screening protocols and treatment monitoring, real time prevention and increased safety. 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 are exposed to external forces;
- 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;
- there is a complex dependence structure between time series;
- there can be unknown and complex causes of observed abnormal behavior;
- there is the 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.

We produce innovative solutions for key problems facing our partners, by developing original statistical and machine learning methodologies.

BigInsight - Statistics for the knowledge economy

PARTNERS

- Norsk Regnesentral (NR) (host institute)
- University of Oslo (UiO)
- University of Bergen (UiB)
- ABB
- DNB
- DNV
- 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 2022, where all partners are represented. In addition to close cooperation with the researchers at NR and the universities, there have been seminar series on broader topics, like Explainable AI, where partners have met and exchanged ideas. This has resulted in bilateral partner-to-partner cooperation across the Innovation Objectives. The bi-weekly BigInsight lunches is another important forum to meet and discuss ideas and solutions across the partners and sectors.

The annual Big Insight Day was successfully arranged at NAV in October with close to 100 participants. The program included numerous short fire talks that gave a very interesting impression of the variety of activities within Big Insight.

ORGANISATION

Board in 2022

Karl Aksel Festø, DNB, chairman
Stian Braastad, ABB
Hans Anton Tvete, DNV
Birgitte F. De Blasio, Folkehelseinstituttet
Erlend Willand-Evensen, Gjensidige
Valentin Johannes Koestner, Hydro
Jon Vegard Sparre, NAV
Lars Holden, Norsk Regnesentral
André Teigland, Norsk Regnesentral
Martin Sending, Oslo University Hospital
Alexander Bjerke, Skatteetaten
Xeni Dimakos, SSB
Kenth Engø-Monsen, Telenor
Bård Støve, University of Bergen
Nadia Slavila Larsen, University of Oslo

Observers: Terje Strand, Research Council of Norway

The board had 2 meetings in 2022. All partners are represented in the Board.

Legal organisation

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

Center Leader

Prof. Arnoldo Frigessi, UiO and OUS, Director

Co-Directors

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

Principal Investigators

Kjersti Aas, NR
Arnoldo Frigessi, UiO
Ingrid Glad, UiO
Clara Cecilie Günther, NR (until 31.08.2022)
Martin Jullum, NR
Alex Lenkoski, NR
Anders Løland, NR
Carlo Mannino, UiO
Hanne Rognebakke, NR
Ida Scheel, UiO
Magne Thoresen, UiO

Administrative Coordinator

Unni Adele Raste, NR

Scientific Advisory Committee (SAC)

Prof. Idris Eckley, Lancaster Univ., UK (chair)
Prof. Samuel Kaski, Aalto University, Finland and University of Manchester, UK
Prof. Geoff Nicholls, Univ. of Oxford, UK
Prof. Marina Vannucci, Rice Univ., Houston, USA
Prof. Veronica Vinciotti, Univ. of Trento, Italy

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

Explaining AI

RESEARCH STRATEGY

We aim to new, interesting, and surprising solutions, which take the field and our partners ahead in their innovation 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. Final transfer to partners' operations happens 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	 Explaning AI
INNOVATION PARTNERS	DNB Gjensidige NAV Skatteetaten Telenor SSB	DNV Kreftregisteret OUS Telenor	DNB Gjensidige Skatteetaten	ABB DNV SSB	Hydro Energy	all partners
RESEARCH PARTNERS	NR UiO UiB	UiO OUS NR NIPH	NR UiO UiB	NR UiO	NR UiO	NR UiO
PRINCIPAL INVESTIGATORS						
Principal Investigators: co-Principal Investigators:	Kjersti Aas Ida Scheel	Magne Thoresen Clara Cecilie Günther until 31.08.2022	Anders Løland Martin Jullum	Ingrid Glad Hanne Rognebakke	Alex Lenkoski Carlo Mannino	Anders Løland Arnoldo Frigessi

METHODS

We solve innovation challenges of our partners by developing new or applying state-of-the-art statistical, mathematical, and machine learning methods in these fields:

- Probabilistic Bayesian models and forecasting
- Complex dependence models
- Latent variable models
- Data integration and knowledge incorporation
- Physics informed machine learning
- Knowledge based machine learning
- Multi-type and multi scale models
- Scalable approximation algorithms
- High dimensional data and time series
- Change and anomaly detection and prediction
- Local and global explanations of black box models
- Networks and graphical models
- Preference learning
- Mechanistic multi-type models in oncology
- Time-to-event models

"As a first step, industrial leaders could gain a better understanding of AI technology and how it can be used to solve specific business problems.

They will then be better positioned to begin experimenting with new applications."

McKinsey: "The future is now: Unlocking the promise of AI in industrials", December 6, 2022

Center leader Arnoldo Frigessi meets Minister of Trade and Industry, Jan Christian Vestre
Foto: Oslo Science City

SCIENTIFIC HIGHLIGHTS

Center Leader and Director
Arnoldo Frigessi

BigInsight has reached impressive achievements during its first seven years: the scale of its translational work, from the development of new methods to the implementation of results at the private and public partners, is rather unique in an international perspective; the range of its activities, across diverse applied areas and industrial sectors, embracing many disparate scientific approaches and techniques, is outstanding. The research production spans from scientific papers published in the best journals to many innovation products and services, which are already in operational use at the partners, or close to be.

BigInsight has a family of private and public partners that motivate, support, contribute and guide the translation of methods into innovation, a type of collaboration which is quite unique and precious. For example, BigInsight has contributed importantly to the successful management of the Covid epidemics in Norway by providing models, methods and results to the National Institute of Public Health (NIPH), also a BigInsight partner. It is special that the work by BigInsight in the years 2018 and 2019 turned out to be fundamental for the Norwegian Covid modelling. The BigInsight NIPH-UiO-NR-Telenor team delivered estimates of the reproduction numbers and predictions of hospital admissions, based on mobile phone mobility data already from March 2020. We believe this work was the first of this type worldwide, and probably still today a rather unique approach in scale and importance.

In all its six innovation areas (Marketing, Health, Sensor, Fraud, Power and Explainability), the research results of BigInsight are very innovative, often based on a combination of statistical method, machine learning approaches and algorithms, and efficient implementation.

The next two final years of BigInsight, 2023 and 2024, will be dedicated to harvesting scientific results, to be published in the best international journals and conferences, and innovation results, to be injected into operations, products and services.

We will also work to build the next BigInsight2.0, as the legacy of our centre has important potentials.

"Unless we root our discipline in problem-solving, a crisis loom"

Leo Breiman (1928-2005), speaking about statistical science

Highlights from our work in 2022

A new method for fraud detection

Due to limited resources, investigators in fraud detection are often restricted to manually control a limited number of suspicious cases (insurance claims, financial transactions, etc.) for potential fraud. The most efficient manner of allocating these resources is to try selecting the k cases with the highest probability of being fraudulent. When fitting models used to estimate the probability of fraud, a loss function is typically optimized. We derive a new loss function, the fraud loss, which is specifically aimed at finding the k largest probabilities. Both simulations and applications to a credit default dataset suggested that the method provided better results compared to alternative methods.

Using Shapley Values and Variational Autoencoders to Explain Predictive Models with Dependent Mixed Features

We are among the teams internationally working on extensions of Shapley values, which are today extensively used to explain complex predictive machine learning models. Precise Shapley value estimates for dependent data rely on accurate modeling of the dependencies between all feature combinations. We propose a method based on variational autoencoder with arbitrary conditioning. The method outperforms the state-of-the-art methods for a wide range of settings. This paper was published in the prestigious Journal of Machine Learning Research.

The mix-lasso model for biomarker detection in cancer

Current statistical models for drug response prediction and biomarker identification fall short in leveraging the shared and unique information from various cancer tissues and multi-omics profiles. We developed the mix-lasso model that captures tissue-specific effects of features on pan-cancer response prediction. The model takes into account both the similarity between drug responses and the heterogeneity between multi-omics data. When applied to large-scale pharmacogenomics dataset from the Cancer Therapeutics Response Portal, mix-lasso enabled accurate drug response predictions and identification of tissue-specific predictive features. Compared to tree lasso model, mix-lasso is more interpretable and stable for drug discovery applications.

Can you tell if a new product will have a viral success in the market?

Understanding the spreading process of new products provides valuable knowledge that can be used for effective marketing. We propose a data-driven agent-based methodology. Inference and predictions are based on short-time observations of the product adoption history and knowledge of the social network of consumers. We model and predict adoptions at the agent level as driven by unobserved peer-to-peer influence and external marketing. Our computationally efficient algorithm is demonstrated on real data from Telenor, predicting the process far into the future using data from a short period after launch.

A new algorithm for mathematical models of cancer treatment in the clinics.

Mathematical modeling is a promising approach to personalized cancer medicine. Yet, the complexity, heterogeneity and multi-scale nature of cancer pose significant computational challenges. Coupling discrete cell-based models with continuous models using hybrid cellular automata (CA) is a powerful approach for mimicking biological complexity. However, for clinically relevant cancer biopsies, such models become computationally very expensive and too slow. We developed parallel algorithms to link stochastic CA with differential equations. We demonstrated our method on a complex hybrid CA model of breast cancer treated with combination chemotherapy. We applied the algorithm to a problem that is 500 times larger than any previous work, allowing us to run personalized therapy simulations based on heterogeneous cell density and tumor perfusion conditions estimated from magnetic resonance imaging data on an unprecedented scale.

Sentiment analysis implemented at Gjensidige

Sentiment analysis uses natural language processing or text analysis to systematically identify, extract, quantify, and study affective states and subjective information. In the case of fraud, certain sentiments, like "impatient" or "unsatisfied", or the transitions between them are possibly seen as a signal of fraudulent behavior. We have developed a method to predict sentiments of insurance chats and tested it at Gjensidige. Chats are instant messages that customers can use to ask questions to customer service. Predicting sentiments is a difficult problem since even humans can disagree on which sentiment(s) can be found in a specific text. The method is implemented at Gjensidige.

Adjustment for seasonal cycle in solar production forecasting

As the amount of renewable energy production continues to increase, forecasts of the total amount of renewable production have become increasingly important for our industrial partners. In particular, solar production in Germany and the Nordic countries has become a critical part of the model stack, especially since it can vary considerably on a daily basis. While we had already developed a model for solar production, when used in the operational setting at Hydro, we noticed that it began behaving strangely in the early and late hours in the days of springtime. We investigated this behavior and discovered this was due to the annual daylight cycle: in the early spring, hours which had previously been completely dark (and thus produced no solar power) began producing solar energy. We adjusted the manner by which our models were trained to account for these discrepancies and helped Hydro implement a refined model.

How much more dangerous is the Omicron SARS-CoV-2 variant?

Understanding the epidemic growth of the novel SARS-CoV-2 Omicron variant was critical for public health. We used unique Norwegian contact tracing data to estimate the differences between Omicron and the previous variant, called Delta. The relative risk of Omicron compared to delta is 1.41 (95% CI 1.27-1.56). We observed increased susceptibility to Omicron infection in household contacts compared to Delta, independent of contacts' vaccination status. We found increased Omicron transmissibility from primary cases to contacts in all vaccination groups, except 1-dose vaccinated, compared to Delta. We concluded that three-dose vaccinated primary cases with Omicron infection can efficiently spread in households, while three-dose vaccinated contacts have a lower risk of being infected by Delta and Omicron. The results were published on Nature Communication and used by Folkehelsainstituttet in the management of the covid pandemics.

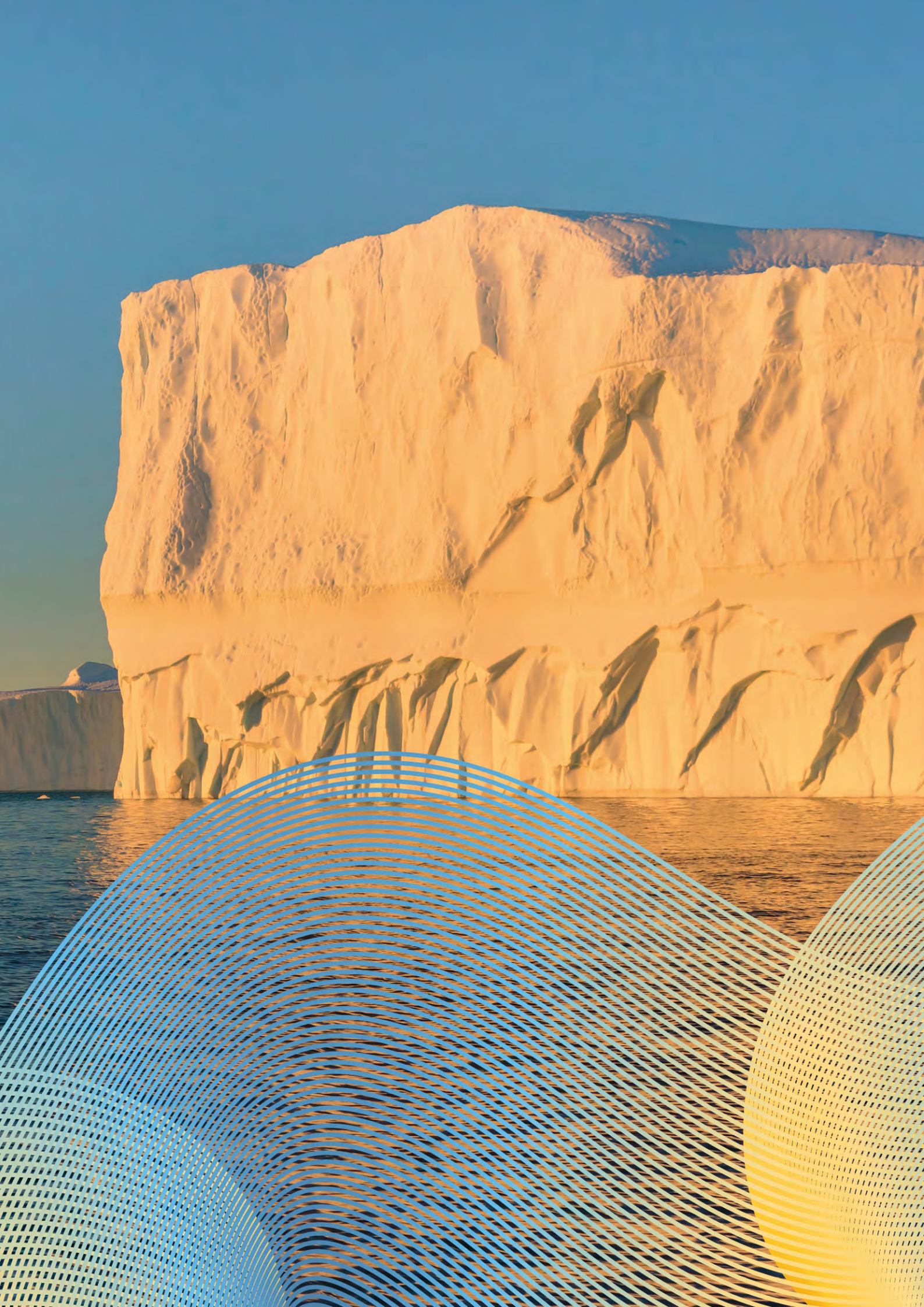
Is your Lithium-ion battery in good health?

Lithium-ion batteries are important for the electrification of the transport sector. The "green shift" is taking place in the marine industry too, where the number of battery-powered vessels is growing. Monitoring the battery State of Health (SoH) is essential in order to optimise battery use, promote safety, and ensure the energy demands. Classification societies like DNV typically require annual capacity tests for this purpose; these tests are disruptive, costly and time-consuming. We propose a novel semi-supervised learning approach to estimate the SoH of a lithium-ion battery system with no labelled data, starting from a minimal set of weakly labelled data from another similar system. Our procedure can be used to significantly progress in failure prevention, operational optimisation, and for planning batteries at the design stage. The method is in use at Corvus.

"Machine learning is the mathematical and computational engine of AI: a fundamental force of technological progress in our increasingly digital, data-driven world.

– But, despite striking advances in the collection and analysis of data, the positive impact of machine learning is not yet sufficiently transformational. In companies, organisations, science and society at large, it has still to show all its potential."

Arnoldo Frigessi, director of BigInsight, 1.3.2023



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 influence the probability of specific actions of the users. We exploit users' behavioural measurements in addition to more traditional characteristics, and external data (including competitors' activity, market indicators, financial information, 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 2022:

Bayesian methodology for recommender systems

BigInsight has over some years developed two new Bayesian approaches to recommendations. The first is based on the Bayesian Mallows Model, and has been shown to perform as well as the state-of-the-art, and in addition providing a higher level of diversity, meaning that users potentially get more interesting, less obvious, recommendations, and the catalogue of offered items is explored more. However, the original methodology had slow convergence, so making it scalable has been of great importance in order for it to be production ready. Our new proposed Variational Bayesian approximation algorithm speeds up convergence drastically. In 2022, we submitted the paper entitled Pseudo-Mallows for Efficient Probabilistic Preference Learning for publication, and published it open on arXiv.

Stochastic models for early prediction of viral customer behaviour on networks

BigInsight has over many years developed methodology for early prediction of the adoption of new products with viral potential on social networks. The proposed stochastic model is agent-based at the individual level and allows for influence both from viral word-of-mouth effects and external factors, such as marketing campaigns. Inference is based on fitting the model via maximum likelihood to data on the early history of adoptions of the product on the social network of the customer. Prediction of the future is performed by simulation with a computationally very efficient algorithm. The method is validated on simulated data, as well as a real telecom dataset provided by Telenor. In 2022 the first paper on this, entitled An agent-based model with social interactions for scalable probabilistic prediction of performance of a new product, was accepted and published in International Journal of Information Management Data Insights. Two follow-up papers are in progress. The first one was in 2022 revised for a new submission, and concerns studying the effect of having a partially observed network on results and conclusions. In the paper we show that if less than half of the links are missing, the results are still qualitatively reasonable. This is important because in practice, the full social networks cannot be fully observed, and in particular, observing all social links is not possible. The second paper following up on the published methodology, for which in 2022 all the theory was developed, as well as experiments and simulations conducted, extends the initial model to also include a preliminary tattling stage. The extension assumes additionally that each vertex connected with an infectious neighbor may also spread influence to their susceptible neighbors, representing tattling. When a vertex turns infectious, it may thus exert influence directly and indirectly, with the latter through the tattlers. The model is shown to be stable and identifiable, and to accurately identify tattling if it is indeed present.

Sales prediction

DNB Bedrift (former DNB Puls) is an app for people running small or medium sized businesses. The app is very popular, currently having 65000 users. Among other things, it produces forecasts of future income based on time series of previous values. In this project, the aim is to improve

the current income predictions. This is a difficult problem, because the historical time series data are very noisy with irregular patterns and many missing values. We use a new combination of traditional time series methods and more recent machine learning methods. The developed code has been transferred to DNB and is implemented in the app.

Explanation of predictions from Black-Box models

In some applications, complex hard-to-interpret machine learning models like deep neural networks are currently outperforming the traditional regression models. Interpretability is crucial when a complex machine learning model is to be applied in areas where trust in the algorithm is required, like for example in clinical applications, fraud detection or credit scoring. There has been a big interest in explaining AI in the context of personalised marketing. In 2022 we have been working with a type of methods denoted counterfactual explanations. More specifically, we have developed MCCE (Monte Carlo sampling of valid and realistic Counterfactual Explanations), a counterfactual explanation method that generates actionable and valid counterfactuals by modelling the joint distribution of the features with an autoregressive generating model where the conditionals are estimated using decision trees. A paper describing the method has been submitted.

Risk impact of weather conditions on car crash counts

In this project we have studied the effect of weather conditions on claim frequencies for motor vehicles. Gjensidige wants to know whether any of the regional differences and/or yearly changes observed for claim frequencies may be attributed to different weather conditions. To investigate this, we use a Generalized Additive Model (GAM) to model the claims. The main conclusion is that including meteorological variables slightly improves predictive performance. The yearly changes are however almost the same before and after including meteorological variables. For the regional differences, the picture is a bit different. When taking temperature and rainfall into account, the eastern part of Norway gets a slightly higher risk, while the western and northern part of Norway gets a slightly lower risk. Lastly, we have evaluated the effect of climate change on the number of car insurance claims. The results show that there will be an average decline in claim frequencies in the future. This is mainly due to the projected increase in temperature and the estimated negative relationship between temperature and the number of claims.

Synthetic data generation

The demand and volume of data containing sensitive information on persons or enterprises have increased significantly over the last years. At the same time, privacy protection principles and regulations impose restrictions on access and use of individual data. Recently, there has been several initiatives to generate synthetic data. If the resulting synthetic data closely resembles the original data, it will make it easier for institutions in the private and public sector to share realistic individual-level micro data while minimizing the risk of disclosing confidential and sensitive information. In this project we have identified the most promising methods for generating synthetic data suggested in the literature, both by investigating the statistical properties and the utility of the methods, and by applying a "motivated intruder" test framework. As far as the latter is concerned, we have focused on what is called differential privacy, investigating how generative models like probabilistic graphical models, diffusion models and normalizing flows may be modified to fulfil the differential privacy criterion.

Automatic index generation

Many of Statistics Norway's traditional surveys are time consuming and labour intensive. There is therefore a need for more efficient, semi-automatic and continuously updated statistics utilizing detailed streams of data and machine learning. To enable this exciting development, we have also in 2022 developed new machine learning methods suited to handle and impute missing values when estimating nutritional values from consumption data. Our work was presented at the 29th Nordic Statistical Meeting, and our methods were supposed to be applied in the forthcoming Survey of consumer expenditure from Statistics Norway, but this has been delayed due to a legal dispute between the Norwegian Data Protection Authority and Statistics Norway. We have also contributed with related methods in Statistics Norway's forthcoming publication of the updated time use survey.

Principal Investigator
Kjersti Aas

co-Principal Investigator
Ida Scheel

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 2022:

Personalized cancer therapies: Modelling cancer drugs sensitivity and synergy in in-vitro screening

The analysis of whole genomes of pan-cancer data sets provides a challenge for researchers, and we have developed a novel rank-based Bayesian clustering method to tackle this unsupervised problem. The advantages of our method are the integration and quantification of all uncertainties related to both the input data and the model, the probabilistic interpretation of final results to allow straightforward assessment of the stability of clusters leading to reliable conclusions, and the transparent biological interpretation of the identified clusters since each cluster is characterized by its top-ranked genomic features. A paper is published.

Cancer pharmacogenomic screens profile cancer cell lines versus many potential anti-cancer drugs to identify new combinations of drugs that have a high probability to work on individual patients. We work with data generated by our partners at Oslo University Hospital and public data to guide therapy based on the statistical prediction of how drugs will behave for individual tumor samples. To improve predictions, we are exploring both structured penalised regression models and structured priors in multivariate Bayesian models to incorporate prior knowledge about the dependence structure between drugs and between multi-omics profiles of cancer cell lines. In particular, we developed the mix-lasso model that introduces an additional sample group penalty term to capture tissue-specific effects of features on pan-cancer response prediction. The mix-lasso model takes into account both the similarity between drug responses (i.e., multi-task learning), and the heterogeneity between multi-omics data (multi-modal learning). When applied to large-scale pharmacogenomics dataset from Cancer Therapeutics Response Portal, mix-lasso enabled accurate drug response predictions and identification of tissue-specific predictive features in the presence of various degrees of missing data, drug-drug correlations, and high-dimensional and correlated genomic and molecular features that often hinder the use of statistical approaches in drug response modeling. Two papers are published.

For combinatorial treatments, prediction of likely synergistic effects is crucial to suggest efficient combinations. We have proposed PIICM, a probabilistic framework for dose-response prediction in high-throughput drug combination datasets. PIICM utilizes a Permutation Invariant version of the Intrinsic Co-regionalization Model for multi-output Gaussian Process regression, to predict dose-response surfaces in untested drug combination experiments. Coupled with an observation model that incorporates experimental uncertainty, PIICM is able to learn from noisily observed cell-viability measurements in settings where the underlying dose-response experiments are of varying quality, utilize different experimental designs, and the resulting training dataset is sparsely observed. One manuscript is submitted for publication.

Healthcare safety management

There is an extreme amount of information available in electronic health records that can be used to learn the behaviour of healthcare institutions, make predictions, guide treatment choices and so on. We have been working on electronic health record data from Akershus University Hospital (AUH) on a project to explore patients' movements

within the hospital and how these may affect the risk of spread of infections. We have been using network models and focused on the evolution of- and differences between networks, in time and space. The project is mostly descriptive and aims to inform decision makers. One paper is published and the PhD student defended her thesis in 2022.

Exploring clonal heterogeneity in blood cancers for personalised treatment

Our goal is to develop a data-driven modelling framework to improve treatment strategies in blood cancers. BigInsight has strong clinical and experimental collaborations in blood cancer at OUS as well as access to unique datasets. One major obstacle to developing personalized medicine is the presence of cellular heterogeneity within the cancer cell population of each patient. This can lead to a common scenario where a therapy initially succeeds at reducing disease burden, but the cancer eventually rebounds due to the outgrowth of a minor but drug-resistant clone. To address this obstacle, we have developed a new method to estimate and quantify the heterogeneity present in each particular cancer. We use available high-throughput drug screening data to infer the subpopulation substructure. Our statistical platform, called DECIPHER, estimates the number of distinct clones present as well as how these clones respond to a specific drug, based on drug screens of patient samples. This information then feeds into evolutionary models of drug response to therapy, to predict the effect of a drug. We use a combination of mathematical modelling and inference for mixture models. Successful implementation of our method will potentially greatly aid in the management of different types of blood cancers, and potentially also solid cancers.

Mathematical models and Bayesian inference in personalised breast cancer therapy

Current personalized cancer treatment is based on bio-markers which allow assigning each patient to a subtype of the disease, for which treatment has been established. Such patient classification represent a first important step away from one-size-fits-all treatment. However, the accuracy of disease classification comes short in the granularity of the personalization: it assigns patients to one of a few classes, within which heterogeneity in response to therapy usually is still very large. In addition, the combinatorial explosive quantity of combinations of cancer drugs, doses and regimens, makes clinical testing impossible. Our strategy for personalised cancer therapy is in silico, based on producing a copy of the patient's tumour in a computer, and to expose this synthetic copy to multiple potential therapies. We show how mechanistic mathematical modelling, patient specific inference and simulation can be used to predict the effect of combination therapies in a breast cancer. The model accounts for complex interactions at the cellular and molecular level and is able of bridging multiple spatial and temporal scales. The model is a combination of ordinary and partial differential equations, cellular automata, and stochastic elements. The model is personalised by estimating multiple parameters from individual patient data, routinely acquired, including histopathology, imaging, and molecular profiling. The results show that mathematical models can be personalized to predict the effect of therapies in each specific patient. The approach is tested with data from breast tumours collected in a recent neoadjuvant clinical phase II trial at OUS. This year, we have been able to develop a numerical algorithm that allows the simulation of a full biopsy, exploiting parallel computing. This study is possibly the first one towards personalized computer simulation of breast cancer treatment incorporating relevant biologically-specific mechanisms and multi-type individual patient data in a mechanistic and multiscale manner: a first step towards virtual treatment comparison.

Principal Investigator
Magne Thoresen

co-Principal Investigator
Clara Cecilie Günther
(until 31.08.2022)

PERSONALISED FRAUD DETECTION

Insurance fraud is expensive, affects insurance prices for all customers, and is therefore important to detect and prevent. Soft fraud, the exaggeration of legitimate claims, is quite diffuse and difficult to spot. A sustainable welfare system requires implementation of effective measures to limit fraud, such as tax avoidance and tax evasion. Furthermore, money laundering is a serious threat to the global economy.

Fraud detection can be seen as a regression/forecasting problem, where fraud (true/false) is the response, possibly with a potential economic loss. In such situations, there are a great number of covariates/features connected to each case, especially if one considers interactions. Further, the data are class imbalanced because the number of investigated fraud cases is generally low compared to the total number of cases. Another challenge is that the data are gathered over time, and that the quality may vary. In addition, only a small subset of the total number of cases is controlled. The objective is then to produce a trustworthy and interpretable probability of fraud for each new case, that can handle structured and unstructured data, including transactions, relational networks, and other available digital records in a privacy responsible setting. Since a small number of cases are controlled, fraud detection can also be seen as anomaly detection.

What we did in 2022:

Network analysis for fraud detection

Fraud often involves more than a single individual. There could be groups of criminals all acting together, or one or more criminals that utilize businesses, financial services, or other (innocent) individuals to perform the fraud. In such settings, network relations play a fundamental role. This is particularly the case for money laundering where both financial transactions, professional roles, and social relations form networks relevant for modeling. Graph neural networks (GNN) is a model class that allows neural network models to be built on top of such graph data structures. Through BigInsight, a part-time master student in the Data Science program at University of Oslo has been working on this topic for a few years. The project used data from DNB to model and detect money laundering with GNNs working on a heterogenous graph consisting of both transactions and professional role networks. In 2022, the thesis was completed with promising results. We also started to write a scientific paper based on and expanding upon this work.

Surveying the field of embeddings

Our network analysis work over the past few years has spurred us to survey the field of embeddings within statistics and machine learning. This has resulted in two survey papers. In the first paper, we take the reader all the way from statistical embeddings, from principal components, via non-linear embeddings, topological embeddings and topological data analysis, to embeddings on networks. The paper has been accepted for publication in the renowned review journal Statistical Science. The second paper builds

on the first by surveying extensions to embeddings of time series and dynamic networks. That paper has been accepted for publication in Journal of Time Series Analysis. We believe that this comprehensive knowledge will inspire further work.

Detecting structuring or smurfing

Structuring is the act of parcelling what would otherwise be a large financial transaction into a series of smaller transactions to avoid scrutiny by regulators and law enforcement. Criminal enterprises may employ several agents ("smurfs") to make the transaction. Structuring appears in money laundering and other financial crimes. Even though this is a known money laundering technique, methods for detecting smurfing are pretty scarce in the scientific literature. We are devising methods to search for and detect smurfing patterns, which can be used as rules directly or as complex features in a machine learning model. In 2022, we finetuned the methodology based on feedback from DNB. Small-scale tests on real DNB customer data showed promising results.

Sentiment analysis for fraud detection

Sentiment analysis is the use of natural language processing (NLP) or text analysis to systematically identify, extract, quantify, and study affective states and subjective information. In the case of fraud, certain sentiments, like "impatient" or "unsatisfied", or the transitions between them could be a signal of fraudulent behavior. We have developed a method to predict sentiments of Gjensidige insurance chats. Chats are instant messages that Gjensidige customers can use to ask questions to customer service. Detecting sentiments is a difficult problem since even humans can disagree on which sentiment(s) that can be found in a specific text. A variant of the method is already being used by Gjensidige and there is interest from other BigInsight partners as well. During the summer of 2022, we also tested adapting NLP models to provide complete textual answers to chat questions from Gjensidige customers, with interesting results.

Copula regression

Traditional regression methods model the conditional probability of fraud given the covariates directly. In copula regression, which is an upcoming field, this conditional model is instead inferred from the joint distribution of the response and the covariates that are constructed with a copula. This allows for a lot of flexibility, especially for the modelling of interactions. However, the existing inference methods for copula regression handle rather low dimensions only. In BigInsight, we have over the past few years been working on new inference methods that are suitable for the dimensions we encounter in fraud detection.

Fraud loss

Statistical fraud detection typically aims at extracting a subset of the most suspicious cases (insurance claims, financial transactions, etc.) for further investigation, since investigators are typically limited to controlling a restricted number k of cases, due to limited resources. The most efficient manner of allocating these resources is then to try selecting the k cases with the highest probability of being fraudulent. Optimizing such a system is not necessarily the same as optimizing to obtain the most accurate probability estimate for all cases and then ranking them. We propose a loss function, denoted the fraud loss, for selecting the model complexity via a tuning parameter. In 2022, we published a paper based on this method in the Journal of Applied Statistics. The paper contains a thorough simulation study showing comparable or better performance when using fraud loss, compared to the traditional approach.

Principal Investigator
Martin Jullum

co-Principal Investigator
Anders Løland

SENSOR SYSTEMS

Sensor data are multidimensional streams of observations from various sensor systems. In this IO we work mainly on sensor systems in the maritime and industrial sector. In addition, we consider the research activity with Statistics Norway as ´sensoring´ society.

For maritime safety surveillance we develop new approaches based on the 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. Our methodology is however very often of generic value. 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. We also rely on other data sources such as AIS data and meteorological data.

What we did in 2022:

Scalable change and anomaly detection

The PhD funded through the dScience center at UiO, has been working on a new method for fast, sparsity-adaptive change point detection. The new method allows detection and localization of an unknown number of changes in the mean-vector sequence of high-dimensional Gaussian vectors. What is very promising, is that we can prove mathematically that the new method successfully localizes all change-points with a near-optimal error rate and under minimal conditions in all sparsity regimes. The method searches for multiple change-points through Seeded Binary Segmentation, and is therefore highly computationally efficient, obtaining a log linear computational complexity. This project is hugely relevant for the partners of this IO and is jointly supervised from NR and UiO. A paper and an R-package are almost finalized.

This PhD student, Per August Moen, and co-supervisor Martin Tveten from NR also won a hackaton on sensor-data from the energy sector arranged by Statkraft – Krafthack 2022.

In 2022 the concept of detection of overheating in electrical propulsion motors, which we have worked together with ABB on earlier, was revisited, in order to perform training of final models to be implemented on board the ships. During this work, it was discovered that minor changes in the configuration of the motor had fundamentally changed the parameters of the model. The system had experienced "concept drift" - changes in the underlying system that affects the machine learning model. This has initiated a new project to investigate if and how concept drift can be handled in data based models in a safety critical application.

Combining AI and expert knowledge for more efficient monitoring

In 2022, the PhD project using both event log files interpreted as multivariate point processes, and continuous sensor data from ABB systems, for failure detection/classification and prediction for these systems, finalized the first paper. The work extended gradient boosting methods to temporal event data, where intensity functions for each event type are learned as flexible functions depending on both "standard" covariates (e.g. sensor data) and the history of events. The main application of this framework is for predicting future events. A second project focusing on the sensor data and in particular identification of which subsets of sensors that trigger specific events, has also been a topic this year, relying more on parametric models and domain expertise in conjunction with data driven methods. This is still work in progress.

A "nærings-PhD" project with ABB is also progressing. It is dedicated to early detection of incipient bearing faults in rolling element bearings. A first paper is published, based on a adaptive division of the vibration signal into a number of frequency bands, time-domain segmentation algorithm and high-resolution maximum likelihood frequency estimation to discover small vibration pulses excited by the defect in the bearing. In 2022, we have been working on a

second paper focusing on improved bearing fault detection by identification of shock pulses originated from the incipient bearing fault and a third paper on identification of different bearing degradation states using prior information from previous measurement points.

A project with data from ABB where we integrate topology drawings of ships' power systems with sensor data and textual log messages in order to automatically detect and classify power blackouts was completed successfully in 2021. The more challenging prediction problem was investigated in 2022, and has continuated internally within ABB. A significant effort on implementing and updating data models in order to implement automated detection and later prediction has been laid down, but the work is not yet completed.

Inferring the effect of marine bio fouling on loss of performance

In 2022, we finalized a scientific paper on the work we have done with DNV on modelling of loss of performance due to bio growth (fouling) on hull and propellers, using Bayesian GAM and INLA. Loss of performance means for example increased fuel consumption. Data come from various ships, with timepoints for hull and propeller washing and a large amount of time series of relevant operational measurements. A detailed report and code has been transferred to DNV at an earlier stage.

Towards zero emission vessels – li-ion battery health diagnostics and prognostics

Our battery sensor data project with DNV and their collaborator Corvus Energy, which is a major producer of maritime batteries operating in several ferries in Norwegian fjords, has made considerable progress in 2022. We develop methods for data driven monitoring of battery health, based on historical data from operating vessels provided by the battery producer. We started with a study of SOH (State of Health) degradation modelling using publicly available data, because there are extremely few measurements of state of health in the operational data. A paper on this, using multivariable fractional polynomials, was published in 2022.

In order to be able to do the same type of SOH degradation modelling on the operational ferry data, we have developed and published a method for extracting pseudo capacities that can be used instead. This is a semi-supervised learning method, that exploits the relationship between the capacity of the battery and the discharge capacity of similar labelled cycles This method has been of great interest to the battery system producer and has caught a lot of attention.

Furthermore, two master theses on battery data analysis have been successfully defended in June 2022. One was conducting a comparative study of the whole spectre of machine learning methods for the SOH degradation with the publicly available data. The other master project was a study of error-in-variable models for total capacity estimation, based on the operational data from Corvus.

Uncertainty quantification for ship emission models

A new project with DNV on quantifying uncertainty in an emission model, was initiated in 2022. Maritime transport is responsible for a large amount of global Greenhouse Gas (GHG) emissions every year. Stricter rules and regulations on limiting GHG emission put pressure on shipping to reduce its emissions, and the industry is seeking to improve the ship efficiency. Modelling of ship fuel consumption and emission is a fundamental input to evaluate the impacts of shipping on environment and climate, and to evaluate the new measures for reducing GHG emissions. DNV has developed the VERDE model, that estimates ship fuel consumption and emission, based on ship hydrodynamical models, information from the Automatic Identification System (AIS), ship parameters and met-ocean data. The VERDE model produces predictions of fuel consumption, that give satisfactory results on aggregated levels like the total fleet. However, the uncertainty in predictions on ship level is large, and we work on giving a measure of uncertainty in these predictions. This is still ongoing work.

Combining data sources with misclassifications, maintaining privacy (SSB)

With partner Statistics Norway (SSB) we work on estimating the proportion of different, non-overlapping, classes in the population (currently employed/unemployed due to data availability, but could be all types of classes) using data from two parties, where both parties have data with misclassifications. The goal is to develop methods that make Statistics Norway better equipped to utilise data from external companies while safeguarding privacy as much as possible. A paper on adjusting misclassification using a second classifier with an external validation sample, was very well published in 2022.

Principal Investigator
Ingrid Glad

co-Principal Investigator
Hanne Rognebakke

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 2022:

Temperature localization for consumption forecasting

A critical component of any market model for the Nordic region is temperature, as Nordic homes are heated with electricity. Throughout Big Insight, we have worked on methodology to improve forecasts of electricity consumption, ultimately replacing a neural network with a more accurate method based on principal component regression, which was completed in 2021. This model operated on Nordic-wide temperature fields and ultimately forecasted the entire system consumption. Our main effort in 2022 was to "locallize" this framework, by looking at the individual regions of the Nordpool market and associating them with temperature sub-fields. We ultimately found that a combination of Nordic level and local level temperature fields gave the best forecast of region-level consumption. This model was subsequently put into production at Hydro.

Medium-term renewable production forecasting

Having an accurate forecast of renewable energy production is an important component of a market model, since high or low levels of production can have outsized impact on electricity prices. In 2021 we developed a ridge regression framework that substantially improved upon previous models for forecasting these quantities based on weather forecasts. After implementation at the industrial partner, it was discovered that performance declined substantially at the 11 to 15 day-ahead level, occasionally leading to unintuitive behaviour. We investigated these outcomes and concluded that weather forecasts at this time horizon must be pooled to offer any acceptable performance. This adjustment to the modelling framework was then implemented at Hydro.

Solar production forecasting in "shoulder" seasons

As noted above, a critical component investigated in the last years of Big Insight was forecasts of renewable energy production. Accurately forecasting solar power production proved particularly difficult in the beginning of spring. This

is because the models that had been developed are trained on recent historical data, typically on an hourly basis. When a given hour in the day goes from being dark to having some illumination, the industrial partner noticed unstable behaviour of the solar production model. We investigated this phenomenon and concluded that it was due to instability in the splines used in the generalized additive model that forecasted solar production. We proposed a pooling mechanism which addressed this issue. The updated model was then implemented at Hydro.

Principal Investigator
Alex Lenkoski

co-Principal Investigator
Carlo Mannino

EXPLAINING AI

At the intersection between artificial intelligence (AI), transparency, privacy and law, there is a need for more research. This IO focuses on explaining AI's black box models and related issues.

AI, statistical models and machine learning methods can often be seen as black boxes to those who construct the model and/or to those who use or are exposed to the methods. This can be due to: a) complicated models, such as deep neural nets, boosted tree models or ensemble models, b) models with many variables/parameters and c) complex 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:

1. Those who construct or use the model should understand how the model works.
2. Those who are exposed to the model should, and sometimes will, have the right to an explanation about a model's behaviour, for example to be able to contest its decision.
3. 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.

Research at BigInsight can challenge 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?

What we did in 2022:

Seminar series

In 2022, we organized three seminars on the themes: "Towards XAI 2.0: From feature interaction relevances to concept-based explanations", "Explaining Artificial Intelligence: Contrastive Explanations for AI Black Boxes and What People Think of Them" and "Hva kan vi lære av NAVs deltagelse i Datatilsynets regulatoriske sandkasse for ansvarlig kunstig intelligens?" (What can we learn from Norwegian Labour and Welfare Administration's participation in the Norwegian Data Protection Authority's Sandbox for responsible artificial intelligence?). Attendance and discussions were very good, and the seminar series continues into 2023.

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. We continue to add new features to our R package – shapr, which now has been made available also in Python. We have continued to improve our Shapley methods further by 1) using variational autoencoders to model dependent, mixed features better (paper published in Journal of Machine Learning Research), and 2) developing methodology to explore and investigate how different clusters of the training data affect the predictions made by any black-box model (published in Data Mining and Knowledge Discovery). In addition to estimating Shapley values, we have devised a new counterfactual method we have called "MCCE", which can be used to efficiently Monte Carlo sample realistic counterfactual explanations (conference paper submitted). Our work in this area has obtained significant attention, positioning us well internationally.

XAI tree: Practical tool for choosing the appropriate explanation method

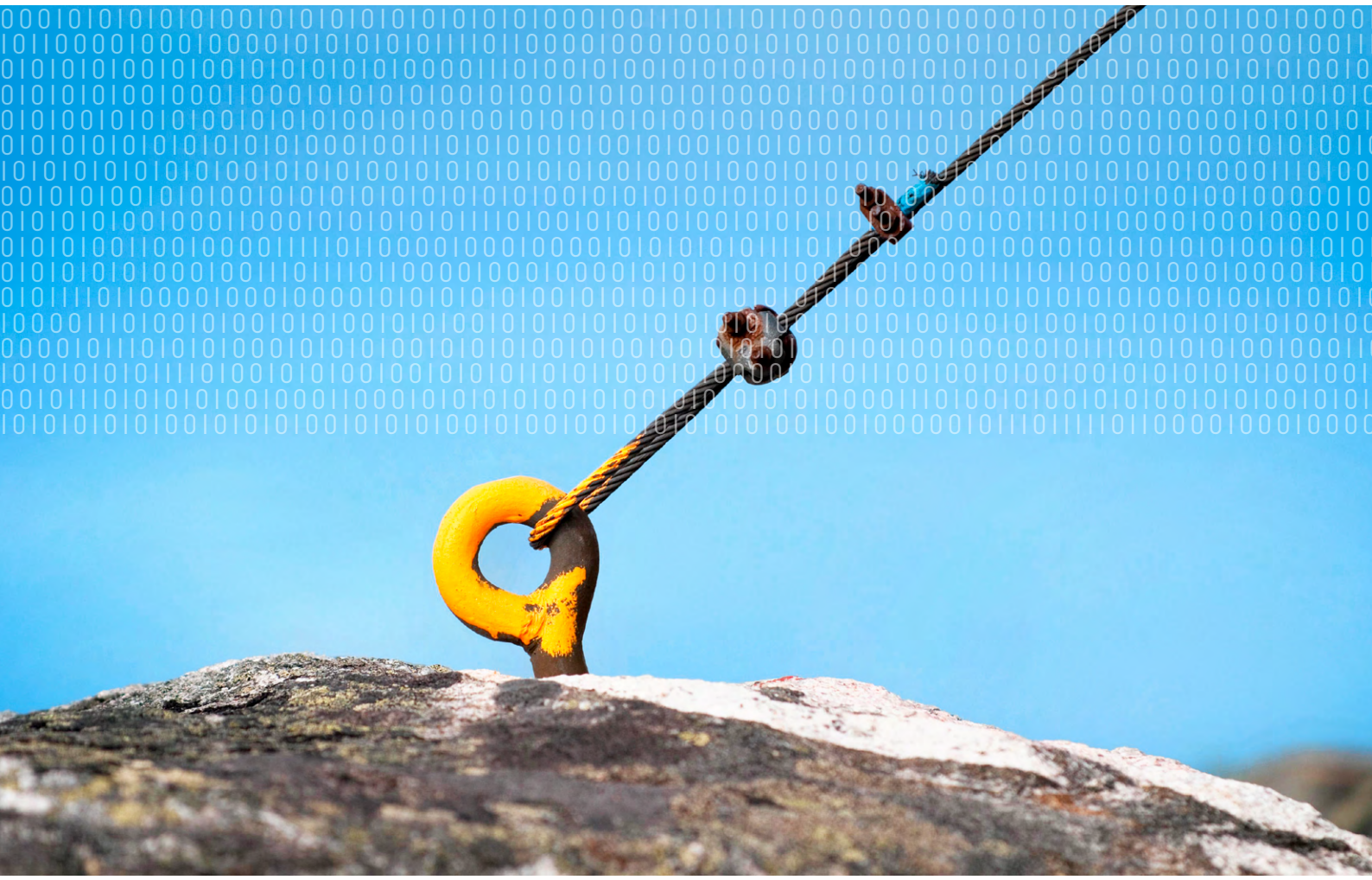
The field of XAI (eXplainable Artificial Intelligence) has in very short time evolved into a myriad of different explanation approaches. Together with NAV we are therefore developing a practical, web based tool, based on answering questions like "Do you want to explain the whole model or specific predictions?". The questions are organised in a tree structure, hence the name "XAI tree". This tool will be used to assist NAV's employees in their quest for responsible artificial intelligence. The tool will also be made available for everyone online.

A national role

We have contributed to various XAI seminars, conferences and the course "Legal Technology: Artificial Intelligence and Law" at the Department of Public and International Law, UiO. NAV has participated with its planned solution for prediction of absence due to illness (and accompanying explanations) in the Norwegian Data Protection Authority's Sandbox for responsible artificial intelligence. We have also contributed to the Norwegian commission for data privacy, which delivered a Norwegian Official Report in September 2022. To summarise: We have been and will continue to be an important voice in the Norwegian AI debate.

Principal Investigator
Anders Løland

co-Principal Investigator
Arnoldo Frigessi



In the era of digitalization, BigInsight produces competence and capacity for the Norwegian knowledge-based economy.

We produce innovative solutions for a sustainable and better society.

BigInsight - Statistics for the knowledge economy

INTERNATIONAL COOPERATION

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

International Academic Partners

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 40 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 chairing each other's scientific advisory boards.

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. BigInsight and the BSU have several joint projects in health and molecular biology. We are both partners in RESCUER, a H2020 project. We will continue our exchange programme, with one Oslo postdoc and one Oslo PhD student working at BSU.

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

Professor Sylvia Richardson, MRC Biostatistics Unit, Cambridge

The Department of Mathematics, University of Minneapolis, USA

This collaboration started in 2018 when Professor Jasmine Foo and associate professor Kevin Leder spent a year at BigInsight, working at the interface between mathematics, cancer biology, clinical oncology, machine learning and statistics. The scientific core of this collaboration is the development of new methods for integrating patient data into mathematical models of cancer, contributing to better treatment for cancer patients. In addition, we develop new educational opportunities in mathematical modelling of cancer at the bachelor's, master's and PhD levels at UiO and UMN. The collaboration is also supported by an INTPART NFR funded project that BigInsight obtained.

University of Hawassa and University of Jimma, Ethiopia

Funded by Norhed, Norpart and the Norwegian Agency for Development Cooperation NORAD, and in partnership with NTNU and UiO, BigInsight concluded in 2020 a ten year project with the University of Hawassa. We now continue our collaboration with these universities, by supervising PhD students. Currently 5 PhD students are working with BigInsight staff, on themes spanning from infectious diseases modelling to diet diversity, from genetic vs environmental causes of non-communicable diseases to tests for hepatitis. We hope to be able to obtain new funding in the future for capacity building of data science in Ethiopia.

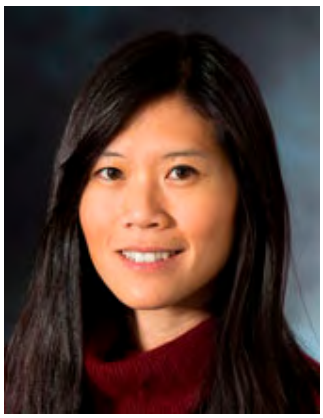
In 2022 we collected and sent many books for teaching and research in statistics and epidemiology to Jimma university.

International guest programme

BigInsight has an international guest programme, funding from short visits up to long-term visiting and adjunct positions. After the covid-19 pandemics, we restarted the programme, and in 2022 we had a one-month visit from Minneapolis.

Students visiting BigInsight

PhD students from other universities spent periods of training and research collaboration at BigInsight. In 2022, we hosted students from ENSTA Paris, University of Milano Bicocca, the Technical University of München and the Deutsches Elektronen-Synchrotron DESY, though the collaboration between NORA and the Helmholtz Information & Data Science Academy.



Jasmine Foo, Deputy Director, Institute for Mathematics and its Applications, Professor, School of Mathematics, University of Minnesota

Kevin Leder, Associate Professor, Industrial and Systems Engineering, University of Minnesota

BigInsight is partner and co-coordinator of the H2020 EU project:

RESCUER: RESISTANCE UNDER COMBINATORIAL TREATMENT IN ER+ AND ER- BREAST CANCER

Breast cancer is the leading cause of cancer-related death in women. Breast cancer is classified into well-recognised molecular subtypes. Despite established molecular classification of tumour subtypes, only some patients benefit from administering drug combinations, which is an indication of tumour heterogeneity. The EU-funded RESCUER project aims to develop a new approach and identify mechanisms of resistance at systems level, exploring how the treatment is affected by patient- and tumour-specific conditions. The project will integrate longitudinal multidimensional data from ongoing clinical trials and novel systems approaches, which combine subcellular/cellular and organ-level in silico models to discover molecular signatures of resistance and predict patient response to combinatorial therapies. This new knowledge will be used to identify already approved drugs with a high curative potential of new personalised drug combinations.

BigInsight is partner of the H2020 EU project:

BD4QoL: Big Data Models and Intelligent tools for Quality of Life monitoring Big Data Models and Intelligent tools for Quality of Life monitoring and participatory empowerment of head and neck cancer survivors. The number of treatment options available for head and neck cancer (HNC) has increased in the last decade thanks to advanced technologies. While current post-treatment care plans focus on functional and health conditions, there are socioeconomic determinants of quality of life that also need to be addressed. The EU-funded BD4QoL project aims to improve HNC survivors' quality of life by developing a person-centred monitoring and follow-up plan. It will use artificial intelligence and Big Data collected from mobile devices, in combination with multi-source clinical and socioeconomic data and patients' reported outcomes. Analysis of the quality of life indicators collected over time will facilitate early detection of risks, prevent long-term effects of treatment, and inform patients and caregivers for personalised interventions.

BigInsight is partner of Transcan EraNet Horizon 2020 project:

ImmuMet: Invigorating immunity against brain metastases in lung and breast cancer patients.

Brain metastases are a devastating complication in breast and lung cancer and a main cause of death. Novel immunotherapies have revolutionized treatment of cancer, but response rates are hard to predict, especially for these patients. In ImmuMet we will characterize identify tumour microenvironment subclasses and unravel their clinical associations. The consortium benefits from the complementary expertise of wet lab scientists with a strong research focus on tumour immunology and brain metastses biology, clinicians who are at the forefront of treatment decisions, partners heading clinical trials and registries for these patients, and biostatisticians who are experts in integrating omics data and developing prediction models.

BigInsight is partner of Nordforsk project:

Data streams and mathematical modelling pipelines to support preparedness and decision making for COVID-19 and future pandemic. The goal of this programme is to, for

Photo: Johannes Jansson -norden.org

the first time, create a joint Nordic long-term academic collaboration on pandemic preparedness using advanced mathematical modelling and systematically collected health data from a broad range of sources. To start off the programme involves Finland (Aalto), Norway (BigInsight), and Sweden (Stockholm), but our ambition is to also include Denmark, Iceland, and the Baltic countries later on. The programme participants comprise epidemiologists, statisticians, mathematicians, and computer scientists. The aim of the programme is to use public health data combined with real-time data streams representing social activity and human mobility, together with advanced mathematical modelling and computational methods to address several of the most urgent questions for COVID-19 and future pandemics: What effects do community structure, individual heterogeneities, and spatial mobility have on reproduction numbers, community immunity, and the efficacy of different preventive measures? How can real-time data streams of social activity and human mobility combined with clinical health data aid in making more accurate predictions and more informed control decisions related to structurally and geographically targeted nonpharmaceutical interventions? How can Nordic health data and novel data streams of relevance for the ongoing COVID-19 and future pandemics be shared and published in a way that allows for better analyses without compromising data privacy of the individuals? The programme will develop methods, tools, and operational procedures for implementing cross-Nordic interoperable public health data pipelines, novel methodology published in international scientific journals, and support the national public health institutes in their aim to keep disease spreading low without causing too high burden on Nordic societies.

Photo: norden.org

Scientific Advisory Committee of BigInsight

Scientific Advisory Committee of BigInsight has five international members. A meeting was organsied in January 2022.

Professor Idris Eckley,
Lancaster University, UK,
Co-Director of the EPSRC-funded STOR-i Centre for Doctoral Training

Professor Samuel Kaski,
Aalto University, Finland, Director of the Finnish Centre of Excellence in Computational Inference Research COIN, and director of the Finnish Center for Artificial Intelligence FCAI.

Professor Geoff Nicholls,
University of Oxford, UK
and St Peter's College Oxford.

Professor Marina Vannucci,
Rice University, Houston, USA,
adjunct faculty member of
the UT M.D. Anderson Cancer Center

Associate Professor
Veronica Vinciotti,
University of Trento, Italy

Research-based innovation, passionately developed in our splendid remote corner of the globe, for global progress and welfare.

BigInsight - Statistics for the knowledge economy

PHD GRADUATES IN 2022

In 2022 the following PhD students affiliated to BigInsight defended their PhD thesis:

Chi Zhang at Institute of Basic Medical Sciences, defended her PhD thesis "Representation and Utilization of Hospital Electronic Health Records Data" Oct. 13, 2022.

Supervisor: Magne Thoresen
Trial lecture: "Machine Learning in the Intensive Care Unit"

Adjudication committee

Associate Professor Cristina Soguero Ruiz, Universidad Rey Juan Carlos
Professor Hugo Lewi Hammer, Oslomet
Professor II Åslaug Helland, University of Oslo

Summary

Electronic Health Records (EHR) data contain the medical and treatment history of patients and have become widely adopted in hospitals in the last decade. Hospital EHR data collected during patient visits contain rich information covering their disease history and progression, medication, procedures, and diagnoses. The availability of large amounts of patient data has brought new opportunities in several research fields, including medicine, epidemiology and method developments using statistical and artificial intelligence tools. Despite the exciting opportunities, using EHR data for research is challenging. The effective extraction and representation of temporal hospital EHR data is a first step to understand the complexity of hospital environment and improve quality of care.

There are two objectives of this thesis. The first objective is to explore different statistical and computational methods to extract, integrate and represent information from temporal and sequential hospital EHR data. In this thesis I explored data mining algorithms (dynamic time warping), machine learning classification algorithms, network analysis on sequential relational data, regression models and regularization, prediction, and variable selection algorithms. The second objective is to demonstrate the broad scope of potential applications of EHR data in the clinical setting. I used two very different hospital EHR datasets (MIMIC-III data from US, AHUS data from Norway) to illustrate the potential applications in patient risk stratification and hospital management and logistic efficiency.

Chi Zhang and supervisor Magne Thoresen

Leiv Tore Salte Rønneberg at Institute of Basic Medical Sciences, defended his PhD thesis "Bayesian modelling of high-throughput drug combination screens: Accounting for uncertainty when searching for drug synergy" on June 20, 2022.

Supervisor: Manuela Zucknick
Trial lecture: "Machine learning for biomarker discovery"

Adjudication committee

Professor Juho Rousu, University of Aalto
Professor Andrea Riebler, Norwegian University of Science and Technology
Professor Marit Inngjerdingen, University of Oslo

Summary

High-throughput drug sensitivity screens in cancer allows large libraries of compounds to be tested in-vitro to determine how a certain cancer responds to various treatment options. When two or more drugs are combined, researchers are frequently interested in finding drug combinations that are synergistic, where the effect of administering drugs simultaneously is greater than administering each drug on their own.

Drug sensitivity screens are noisy by nature, making the exact quantification of synergistic drug effects difficult, and potentially misleading if the noise sources are not accounted for. In this thesis, tools for working with large drug combination datasets are developed that account for the inherent uncertainty in the data. Specifically, the tools can be used for:

- Quantification of synergistic effects and synergy scores, through a Bayesian semi-parametric model that accurately reports the uncertainty inherent in the estimates.
- Discovery of biomarkers for drug synergy, connecting molecular characteristics of cancer cells to drug synergy.
- Predicting dose-response and synergistic effects in unperformed experiments, using incomplete and noisy training data from previous experiments

Master theses at BigInsight

1. Daniel Piacek (2017): Detecting fraud using information from social networks [UiO – Fraud]

2. Jonas Fredrik Schenkel (2017): Collaborative Filtering for Implicit Feedback: Investigating how to improve NRK TV's recommender system by including context [UiO -Marketing]

3. Martin Tveten (2017): Multi-Stream Sequential Change Detection - Using Sparsity and Dimension Reduction [UiO - Sensor]

4. Kristin B. Bakka (2018): Changepoint model selection in Gaussian data by maximization of approximate Bayes Factors with the Pruned Exact Linear Time algorithm [NTNU - Sensor]

5. Simon Boge Brant (2018): Dynamic survival prediction for high-dimensional data [UiO - Health]

6. Tristan Hugh Curteis (2018): Focused Model Selection for Longitudinal Data [UiO - Health]

7. Amanda Haugnes Rygg (2019): GLM and GAM modelling of life insurance data [UiO - Marketing]

8. Jenine Gaspar Corrales (2019): Analyzing and Predicting Demographics of NRK's Digital Users [UiO - Marketing]

9. Eirik Halsteinslid (2019): Addressing collinearity and class imbalance in logistic regression for statistical fraud detection [UiO - Fraud]

10. Amirhossein Kazami (2019): A Semi-Supervised Approach to the Application of Sensor-based Change-Point Detection for Failure Prediction in Industrial Instruments [NTNU - Sensor]

11. Camilla Lingjærde (2019): Tailored Graphical Lasso for Data Integration in Gene Network Reconstruction [UiO - Health]

12. Vegard Stikbakke (2019): A boosting algorithm to extend first-hitting-time models to a high-dimensional survival setting [UiO - Health]

13. Oda Johanne Kristensen (2020): Scalable Markov Chain Monte Carlo by subsampling methods [UiO - Sensor]

14. Vera Haugen Kvisgaard (2020): Can undersampling boost fraud detection? Combining undersampling with stochastic gradient boosting for high-dimensional prediction of rare events [UiO - Fraud]

15. Elyas Dawod Mohammed (2020): Count time series with application to corporate defaults [UiB - Fraud]

16. Aleksander Njøs (2020): Multiple imputation for Cox regression with sampled cohort data [UiO - Health]

17. Lars H.B. Olsen (2020): Likelihood-Based Boosting: Approximate Confidence Bands and Intervals for Generalized Additive Models [UiO – Explain]

18. Hanne Tresselt (2020): Modelling Car Insurance Data with Individual Effects [UiO - Marketing]

19. Fredrik Wollbraaten (2020): Sequential Monte Carlo and twisted state space models [UiO - Sensor]

20. Edward F. Bull (2021): Introducing an Efficient Approach for Expressing Uncertainty in Deep Learning with Bayesian Neural Networks [UiO - Sensor]

21. Håkon Bliksås Carlsen (2021): Studying the application of semi-supervised learning for fraud detection [UiO – Fraud]

22. Bob Betuin Fjellheim (2021): Bayesian Plackett-Luce Models for Describing Consensus in Ranking Data - Review and applications to real data [UiO - Marketing]

23. He Gu (2021): Recurrent Neural Networks for predicting ship motor temperatures aiming to help prevent motor overheating [UiO - Sensor]

24. Nikola Kaletka (2021): Effects of prior information on monotonicity directions in additive monotone regression [UiO - Health]

25. Nicolay Bjørlo Kristensen (2021): Weakly Supervised Learning for Predictive Maintenance: An Extended Random Forest Approach using Imbalanced Event Data from Hybrid Ships [UiO - Sensor]

26. Anna Skovbæk Mortensen (2021): Fraud detection using copula regression [UiO - Fraud]

27. Øystein Skauli (2021): Modelling Short Term Changes in User Interest for Online Marketplaces [UiO - Marketing]

28. Peder Nørving Viken (2021): Sequential Monte Carlo and twisted state space models - Twisting models to reduce variance [UiO - Sensor]

29. Erik Holst Aasland (2022): Shapley values for dependent features using divisive clustering [NTNU – Explain]

30. Christian Grindheim (2022): Comparative study of machine learning methods for battery state of health estimation [UiO - Sensor]

31. Fredrik Johannessen (2022): Finding Money Launderers Using Heterogeneous Graph Neural Networks [UiO – Fraud]

32. Meghana Kamineni (2022): The effect of non-pharmaceutical interventions to control mobility during the Covid-19 pandemics [UiO - Health]

33. Anna Kejvalova (2022): Total capacity estimation for marine batteries using measurement error regression [UiO - Sensor]

34. Shuijing Liao (2022): The application of penalized logistic regression for fraud detection Studying measures of prediction performance for class imbalanced and high-dimensional data [UiO - Fraud]

35. Arne Rustad (2022): Model-based counterfactual explanations using tabular GAN or VAE [NTNU – Explain]

Active students

36. Jørn Frøysa Hole (planned finished 2023): Handling class imbalance within fraud detection using generative models [UiO – Fraud]

37. Anders Kielland (planned finished 2023): A comparison of machine learning and mechanistic models for predicting treatment response with combined aromatase and CDK4/6 inhibitors in breast cancer patients [UiO - Health]

38. Haakon Muggerud (planned finished 2023): Recommender systems for non-web-based commerce [UiO - Marketing]

39. Thomas Mullaly (planned finished 2023): Data science within insurance [UiB - Fraud]

40. Ingvild Riiser (planned finished 2023): Generating Synthetic Event Data for Labour and Welfare Studies [UiO - Explaining AI]

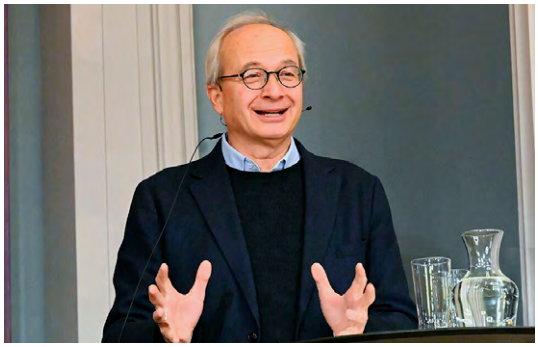
41. Eirik Sjåvik (planned finished 2023): Energy Demand Forecasting [UiO – Power]

ACTIVITIES AND EVENTS

2022 BigInsight Day

The yearly meeting of all researchers and partners of BigInsight was held November 21st at NAV. This is the occasion for the larger BigInsight community to meet, and this was the first physical arrangement in two years. Thanks to all who participated with excellent 6-10 minutes "fire talks", in-depth presentations and a debate about jobs and job market for data-scientists.

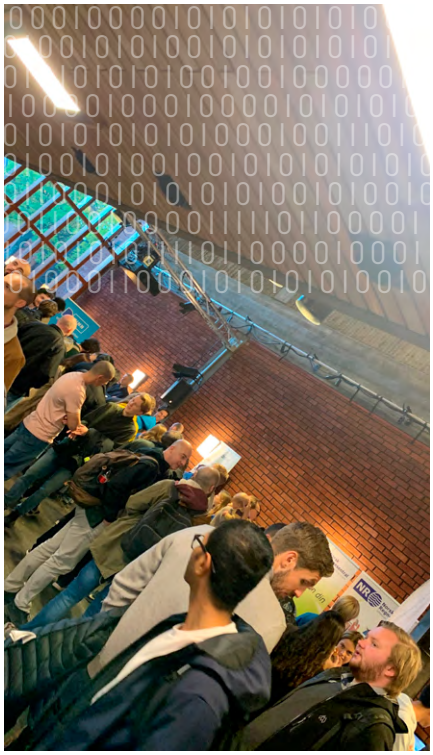
The yearly surprise was the seminar of Miguel Ramalho, data scientist of Bellingcat, who we invited to Oslo. Bellingcat is an international investigative journalism group that specialises in fact-checking and open-source intelligence. Bellingcat had just helped to identify the Russian spy Mikhail Mikushin, who was a guest researcher in Tromsø. Miguel gave a wonderful introduction to the methods used by him in order to fact check events and documents.



Guest speaker Miguel Ramalho from Belllingcat

2022 Oslo Data Science Day

For the fifth time, the Data Science Day was held on October 19th at the University of Oslo. The event had interesting talks from Spotify, UiO and the Alan Turing Institute, mingling with food and drinks, and ca. 15 company stands. Many of the BigInsight partners had stands in order to present themselves as exciting working places. Data Science Day at UiO is organised by dScience, the Centre for Computational and Data Science.



Seminars

BigInsight's biweekly Wednesday lunch takes place at the Department of Mathematics and NR alternatingly. In 2022 15 lunches were organized, see our webside for a list of invited speakers. Only two of the lunches were, due to covid-19 restrictions, held via Zoom.

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.

The Tuesday statistics seminar at the Department of Mathematics, co-sponsored by BigInsight, is a traditional semi-weekly seminar for the whole statistics community in the Oslo area.

TRAINING AND COURSES

The University of Oslo runs the Master Program in Data Science since 2018. Admission to this Data Science master program requires a bachelor with at least two statistics and two computer science courses, plus a solid mathematical foundation, and as such it is different from many other competing programs in Norway, which do not have such requirements. The focus of the master courses is on methods, algorithms, and data analysis pipelines, with less time spent 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. Every year, there is a very large number of applicants of which around 25 are admitted each year. BigInsight participates to the master program by teaching, master projects and industrial contacts. We also contribute to the UiO Honours-programme (bachelor) with some teaching.

BigInsight staff supervise MSc projects in data science and statistics. When possible, we couple these projects to an on-going PhD project, so that the PhD student can participate to the supervision. See list above.

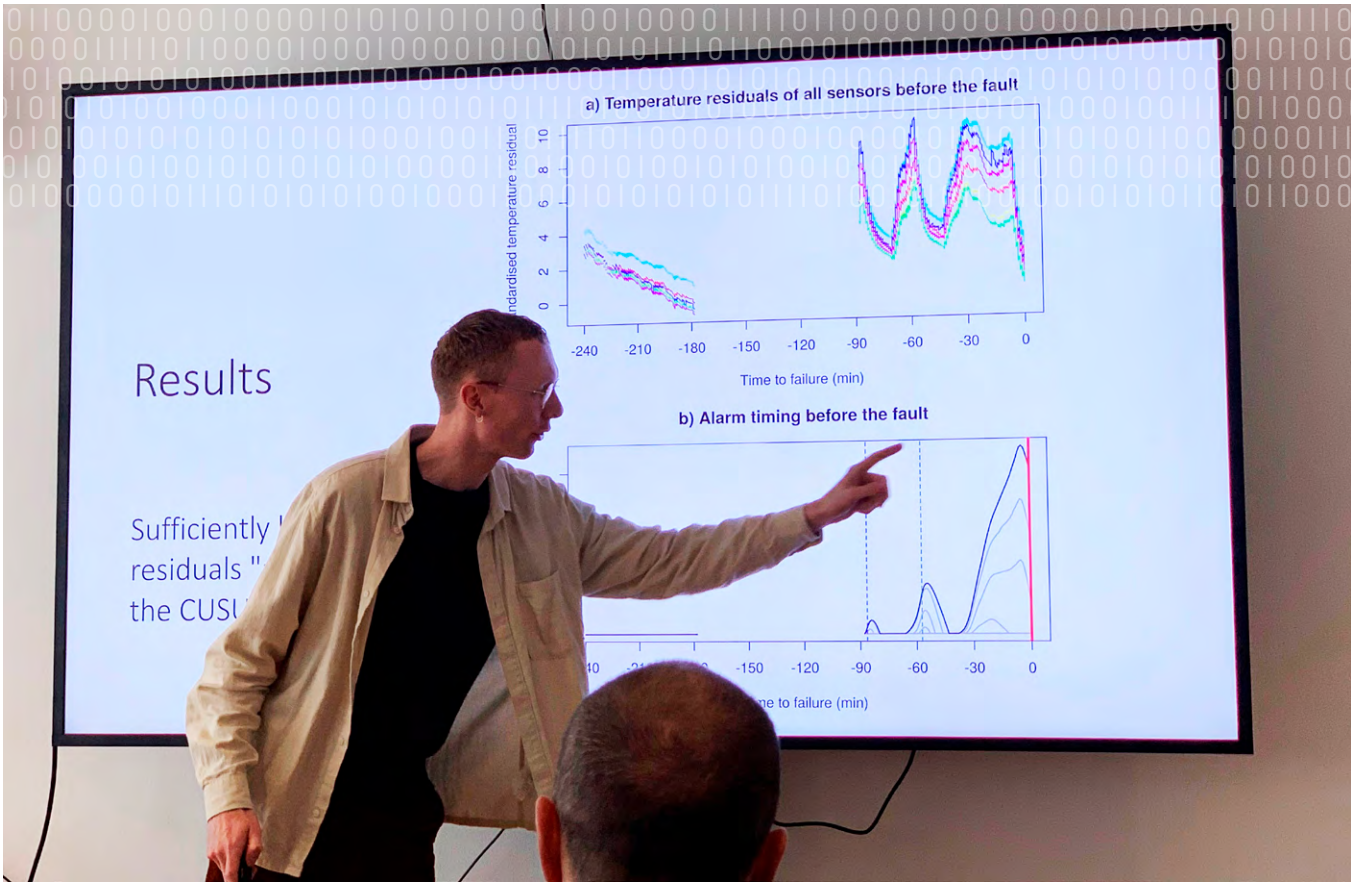
Some PhD students work as teaching assistants, and in the final year also as teachers, in our courses, also at the Faculty of Medicine. Postdocs have teaching duties occasionally and participate in supervision of master and bachelor students.

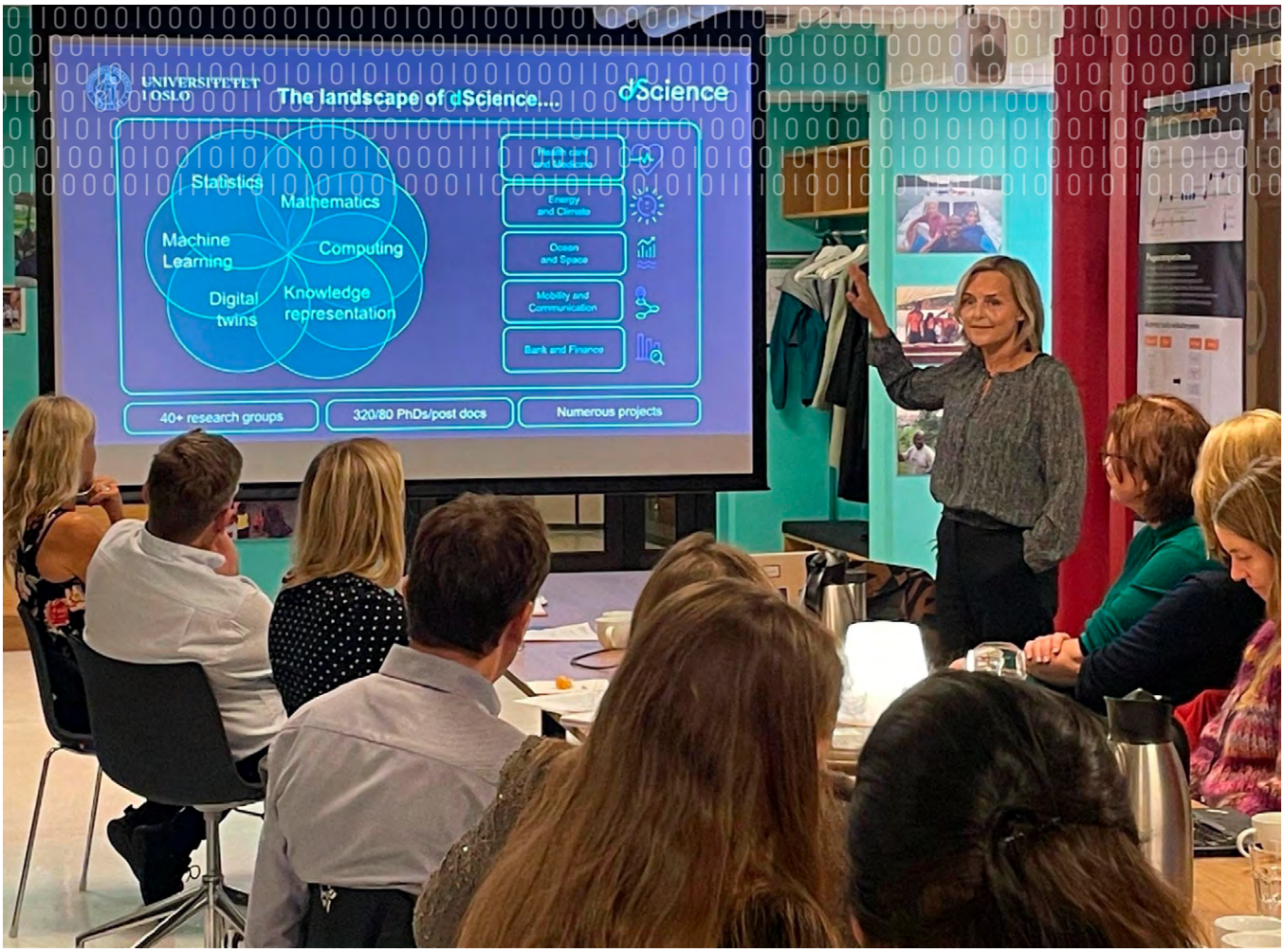
Many PhD students contribute to the advising services in statistics, biostatistics, bioinformatics, and data science, which we offer to researchers at UiO and OUS. They follow an experienced advisor, before they advise on their own (with behind the scene support if needed). We offer a drop-in advising service and a more long term support. In this latter case, students are often coauthors of a research paper. These are very precious experiences. PhD students at OCBE 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. Co-supervision of BigInsight master students together with university staff is also excellent training for young researchers at NR.

New course in mathematical optimization and learning

Mathematical optimization and learning are strongly intertwined. On one hand, optimization algorithms are the indispensable building blocks of learning methods. On the other hand, learning strategies are increasingly exploited to enhance classical optimization algorithms. These are the two main legs of AI systems: learning allows us to describe and predict data (resp. descriptive and predictive analytics), whereas optimization algorithms use the processed data to make the best possible decisions out of it (prescriptive analytics). BigInsight and the Department of Mathematics organised a new course on optimization and its application in learning with the intention of introducing students to mathematical optimization and provide them a coherent basis for further study. The course covered linear and mixed integer programming, combinatorial optimization, flows and minimum cuts, convexity theory, and a pinch of complexity theory. The methods were applied in learning and classification models, such as neural networks and classification trees. Professor Carlo Mannino was responsible for the course.


Results
a) Temperature residuals of all sensors before the fault
Time to failure (min)
b) Alarm timing before the fault


UNIVERSITETET I OSLO
The landscape of dScience...
dScience
Statistics
Mathematics
Machine Learning
Computing
Digital twins
Knowledge representation
Energy and Climate
Ocean and Space
Mobility and Communication
Bank and Finance
40+ research groups
320/80 PhDs/post docs
Numerous projects

COMMUNICATION AND DISSEMINATION ACTIVITIES

BigInsight presented to Minister of Trade and Industry Jan Christian Vestre

Oslo Science City met with Næringsminister Vestre on 8 June 2022. Among the presentation given to the Minister, illustrating the link between research and innovation, BigInsight director Frigessi presented our centre. "As the first "AI" centre in Norway, we work with private and public partners by developing methods and models which give value for the partner" explained Frigessi. Frigessi mentioned our work with Hydro, DNV, DNB, and FHI. Vestre was interested and mentioned that he wants to learn more about BigInsight as an important example of research based innovation.

Website

The website of the center is biginsight.no.

BigInsight outreach presentations

BigInsight has been present in the Norwegian media in 2022 because of our work in Covid-19 modelling and in explainable AI. We hold seminars and participate to the public debate about AI and digitalization. We maintain a list of our public appearances on our webpage. BigInsight participates, through UiO and NR, to the Norwegian Artificial Intelligence Research Consortium (NORA) and to the Norwegian Open AI Lab. We are also central to the UiO centre dScience and play an advising role in the building of the first Norwegian research district Oslo Science City. We were present in Arendalsuka with several speakers.

Foto: Oslo Science City

RECOGNITIONS

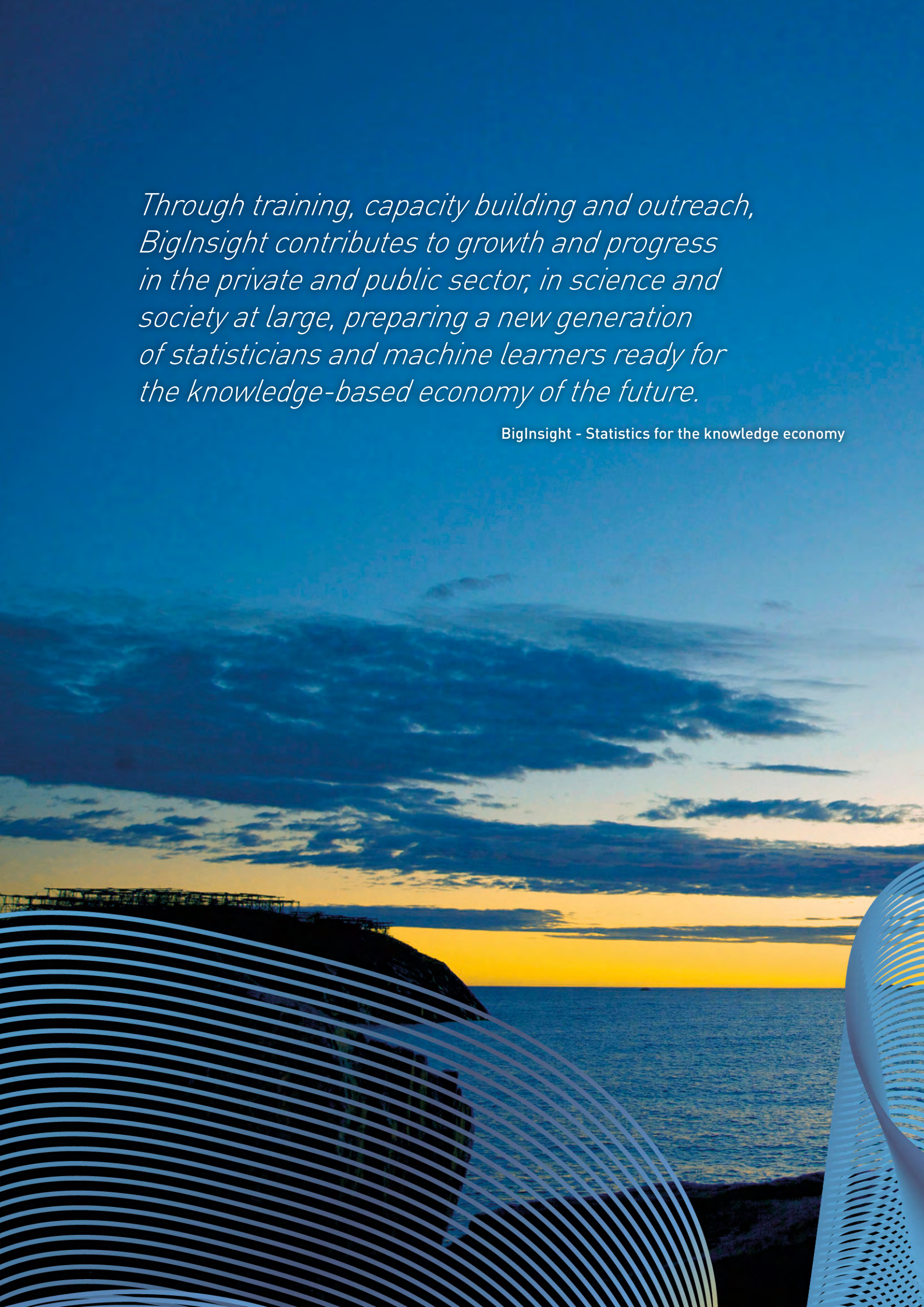
Per August Moen and Martin Tveten won Krafthack 2022.

Statkraft, Energi Norge and Unifai organized Krafthack 2022. The task was to analyse and predict the effect and status of a hydropower plant located in Kvilldal. There, over 3,000 GWh of energy is produced each year using four turbines. These turbines are subject to an ever-changing environment with complex load requirements, as they must start, stop and adjust to varying water conditions. These changing loads result in stresses on the structure, and it is important to monitor the turbine in order to minimize forces on its components and maintain a high level of safety. At Krafthack 2022 many teams competed: by combining data, engineering knowledge and machine learning, new and valuable insights should be gained from Norway's largest hydropower plant. With shared first place, the BigInsight duo Moen and Tveten outperformed the other participating teams. The prize was NOK 25,000. "We struggled a bit at the start to get control of 2 million observations. Then it was of course exiting that our predictions were the best in the end, said Moen to the reporter from dScience.

Per August Moen og Martin Tveten while working at Krafthack 2022 organised by Statkraft, Energi Norge and Unifai. Foto: Statkraft

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.

BigInsight - Statistics for the knowledge economy



PERSONNEL

NAME	INSTITUTION	MAIN RESEARCH AREA
Arnoldo Frigessi	UiO/OUS/NR	Marketing, Health, Sensor, Explaining AI
Stian Braastad	ABB	Sensor
Børre Gundersen	ABB	Sensor
Petter Häusler	ABB	Sensor
Jaroslaw Nowak	ABB	Sensor
Morten Stakkeland	ABB	Sensor
Stian Torkildsen	ABB	Sensor
Andree Underthun	ABB	Sensor
Daniel Warda	ABB	Sensor
Frank Wendt	ABB	Sensor
Per-Espen Aarseth	ABB	Sensor
Jan Nygård	CRN	Health
Qinghua Liu	Oda	Marketing
Lars Erik Bolstad	DNB	Fraud
Bjørn Ingeberg Fesche	DNB	Fraud
Karl Aksel Festø	DNB	Marketing
Fredrik Johannessen	DNB	Fraud
Tobias Lillekvelland	DNB	Marketing
Tatiana Pekarskaya	DNB	Marketing
Daniel Piacek	DNB	Fraud
Aiko Yamashita	DNB	Marketing, Explaining AI
Lars Holterud Aarsnes	DNV	Sensor
Øystein Alnes	DNV	Sensor
Ole Christian Astrup	DNV	Sensor
Håvard Nordtveit Austefjord	DNV	Sensor
Andreas Brandsæter	DNV	Sensor
Hendrik Brinks	DNV	Sensor
Christos Chryssakis	DNV	Sensor
Øystein Engelhardtsen	DNV	Sensor
Ørjan Fredriksen	DNV	Sensor
Odin Gramstad	DNV	Sensor
Bingjie Guo	DNV	Sensor
Qin Liang	DNV	Sensor
Tom Arne Pedersen	DNV	Sensor
Marek Skolimowski	DNV	Sensor
Gaute Storhaug	DNV	Sensor
Hans Anton Tvete	DNV	Sensor
Bjørn-Johan Vartdal	DNV	Sensor
Erik Vanem	DNV	Sensor, Power
Nikos Violaris	DNV	Sensor
Sindre Froyn	Gjensidige	Marketing
Jørgen Andreas Hagen	Gjensidige	Marketing

NAME	INSTITUTION	MAIN RESEARCH AREA
Martin Lerudgjordet	Gjensidige	Marketing
Eirik L. Halsteinslid	Gjensidige	Marketing
Hayat Mohammed	Gjensidige	Marketing
Gunnhildur Steinbakk	Gjensidige	Fraud
Bård Storvik	Gjensidige	Marketing
Daniel Svensson	Gjensidige	Fraud
Geir Thomassen	Gjensidige	Fraud
Stefan Erath	Hydro	Power
Valentin Koestler	Hydro	Power
Ellen Paaske	Hydro	Power
Peter Szederjesi	Hydro	Power
Birgitte De Blasio	NIPH	Health
Jonas Lindstrøm	NIPH	Health
Jørgen E. Midtbø	NIPH	Health
Alfonso Palomares	NIPH	Health
Gunnar Rø	NIPH	Health
Camilla Stoltenberg	NIPH	Health
Robindra Prabhu	NAV	Explaining AI
Cathrine Pihl Lyngstad	NAV	Explaining AI
Jacob Sjødin	NAV	Explaining AI
Lars Sutterud	NAV	Explaining AI
Jon Vegard Sparre	NAV	Explaining AI
Kjersti Aas	NR	Marketing, Explaining AI
Magne Aldrin	NR	Sensor
Solveig Engebretsen	NR	Health
Clara-Cecilie Günther	NR	Marketing, Health (until 31.08.22)
Ola Haug	NR	Marketing, Sensor
Kristoffer Herland Hellton	NR	Marketing, Sensor
Lars Holden	NR	Health, Fraud
Marit Holden	NR	Health
Ragnar Bang Huseby	NR	Power
Martin Jullum	NR	Marketing, Fraud, Explaining AI
Alex Lenkoski	NR	Power
Pierre Lison	NR	Fraud
Anders Løland	NR	Fraud, Power, Explaining AI
Linda R. Neef	NR	Marketing
Didrik Nilsen	NR	Explaining AI
Ildikó Pilán	NR	Fraud
Annabelle Redelmeier	NR	Marketing, Explaining AI (until 30.09.22)
Hanne Rognebakke	NR	Marketing, Sensor
André Teigland	NR	Explaining AI
Ingunn Fride Tvete	NR	Health
Martin Tveten	NR	Sensor
Jens Christian Wahl	NR	Marketing, Power (until 31.08.22)
Mette Langaas	NR/NTNU	Sensor, Health
Tero Aittokallio	OUS	Health

NAME	INSTITUTION	MAIN RESEARCH AREA
Pilar A. Duran	OUS	Health
Torsten Eken	OUS	Health
Jorrit Enserink	OUS	Health
Thomas Fleischer	OUS	Health
Maria Serena Giliberto	OUS	Health
Eivind Hovig	OUS	Health
Vessela Kristensen	OUS	Health
Marissa LeBlanc	OUS	Health
Sygve Nakken	OUS	Health
Andrew Reiner	OUS	Marketing, Health
Fredrik Schjesvold	OUS	Health
David Swanson	OUS	Health
Kjetil Tasken	OUS	Health
Anders Berset	Skatteetaten	Marketing, Fraud
Wenche Celiussen	Skatteetaten	Fraud
Rasmus Sjøholt Engelschiøn	Skatteetaten	Fraud
Øystein Olsen	Skatteetaten	Marketing
Nils Gaute Voll	Skatteetaten	Fraud
Jorid Ødegård	Skatteetaten	Fraud
Kim Benjamin Boué	SSB	Marketing, Sensor, Power
Xeni Dimakos	SSB	Marketing
Boriska Toth	SSB	Marketing
Øyvind Langsrud	SSB	Sensor, Power
Li-Chun Zhang	SSB	Marketing, Sensor, Power
Kenth Engo-Monsen	Telenor	Marketing, Health, Sensor
Dag Tjøstheim	UiB/NR	Fraud
Elja Arjas	UiO	Marketing
Jukka Corander	UiO	Health
Ingrid Dæhlen	UiO	Fraud
Riccardo de Bin	UiO	Marketing, Fraud
Ingrid K. Glad	UiO	Sensor
Ingrid Hobæk Haff	UiO	Fraud
Nils Lid Hjort	UiO	Sensor, Fraud
Carlo Mannino	UiO	Power
Waldir Leoncio Netto	UiO	Marketing, Health
Ida Scheel	UiO	Marketing
Geir Storvik	UiO	Sensor
Øystein Sørensen	UiO	Health
Magne Thoresen	UiO	Health
Valeria Vitelli	UiO	Marketing, Health
Manuela Zucknick	UiO	Health

NAME	FUNDING	NATIONALITY	PERIOD	GENDER	TOPIC
Postdoctoral researchers with financial support from BigInsight					
Haakon C. Bakka		Norway	2020-2022	M	Sensor
Annika Krutto		Estonia	2020-2023	F	Health
Alvaro Köhn Luque		Spain	2021-2023	M	Health
Postdoctoral researchers in BigInsight with financial support from other sources					
Tugba Akman	Turkey Research Council	Turkey	2021-2022	F	Health
Theophilus Quachie Asenso	UiO	Ghana	2021-2023	M	Health
Youness Azimzade	UiO	Iran	2021-2023	M	Health
Hilde Kjelgaard Brustad	UiO	Norway	2022-2024	F	Health
Erlend Ignacio Fleck Fossen	UiO	Norway	2022-2024	M	Health
Neda Jalali	NIPH	Iran	2021-2022	F	Health
Fatih Kizilaslaw	UiO	Turkey	2021-2023	M	Health
Richard Xiaoran Lai	UiO	UK	2019-2022	M	Health
Euloge Clovis Pagui	UiO	Italy	2022-2024	M	Health
Henry Pesonen	UiO	Finland	2019-2022	M	Health
Vincenzo Politi	UiO	Italy	2021-2022	M	Health
Vandana Ravindran	UiO/IMB	India	2020-2023	F	Health
Leonardo Santana	UiO	Brasil	2020-2023	M	Health
Leonard Schmiester	UiO	Germany	2021-2023	M	Health
Mauricio M. Soares	UiO	Brasil	2020-2023	M	Health
George Zhi Zhao	OUS	China	2021-2023	M	Health
PhD students with financial support from BigInsight					
Ingrid Dæhlen		Norway	2021-2024	F	Several
Riccardo Parviero		Italy	2018-2022	M	Marketing
Clara Bertinelli Salucci		Italy	2019-2022	F	Sensor
Jonas Fredrik Schenkel		Norway	2018-2022	M	Sensor
Fredrik Wollbraaten		Norway	2020-2023	M	Sensor

NAME	FUNDING	NATIONALITY	PERIOD	GENDER	TOPIC
PhD students in BigInsight with financial support from other sources					
Henok Asefa	Norad	Ethiopia	2021-2023	M	Health
Sebastian Bieringer	NORA	Germany	2022	M	Health
Andrea Bratsberg	UiO	Norway	2021-2023	F	Health
Simen Eide	Finn.no, NæringslivPhD	Norway	2018-2023	M	Marketing
Emanuele Gramuglia	ABB	Italy	2016-2022	M	Sensor
Elisabeth Griesbauer	UiO	German	2022-2024	F	Health
Lars Petter Johnsen	UiO	Norway	2021-2023	M	Explaining AI
Teshome Kabeta	Norad	Ethiopia	2021-2023	M	Health
Per August Moen	dScience	Norway	2021-2025	M	Sensor
Even Moa Myklebust	NFR	Norway	2020-2023	M	Health
Jaroslaw Nowak	ABB, NæringslivPhD	Poland	2018-2023	M	Sensor
Lars H. B. Olsen	MatNat/UiO	Norway	2020-2024	M	Explaining AI
Anja Stein	STORi, Lancaster	Norway	2019-2023	F	Marketing
Andrea Chi Zhang	NIPH	China	2016-2022	F	Health
Emilie Ødegård	UiO/IMB	Norway	2019-2023	F	Health, Marketing
Master degrees					
Henning Smedsrud Aldrin	2022-2024	M	Fraud	M	Fraud
Amir Basic	2022-2024	M	Marketing	M	Marketing
Linyinfeng Deng	2022-2024	M	Sensor	M	Sensor
Hanne Gjemdal	2022-2024	F	Fraud	M	Sensor
Jonas Quesada Gjessen	2022-2024	M	Health	M	Fraud
Elisabeth Griesbauer	2022-2024	F	Health	F	Power
Christian Grindheim	2020-2022	M	Sensor	F	Health
Jørn Frøysa Hole	2021-2023	M	Fraud	F	Health
Meghana Kamineni	2021-2022	F	Health	F	Sensor
Anna Kejvalova	2020-2022	F	Sensor	M	Health
Anders Kielland	2021-2023	M	Health	M	Sensor
Haakon Muggerud	2021-2023	M	Marketing	F	Fraud
Mamun Ar Rashid	2022-2024	M	Sensor	F	Fraud
Ingvild Riiser	2021-2023	F	Fraud	M	Marketing
Eyvind Skretting	2022-2024	M	Sensor	F	Fraud
Jonas Einar Thorsen	2021-2023	M	Fraud	M	Marketing
Han Yu	2022-2023	F	Sensor	M	Fraud
Research assistant (vitas)					
Severin Schirmer			2021-2022	F	Health

FINANCIAL OVERVIEW

FUNDING	1000 NOK
The Research Council	6 780
Norwegian Computing Center (NR)	924
Research Partners*, in kind	9 652
Research Partners*, in cash	1 500
Enterprise partners**, in kind	1 529
Enterprise partners**, in cash	4 545
Public partners***, in kind	2 399
Public partners***, in cash	1 990
Sum	**29 318**

COSTS	
NR, research	9 595
NR, direct costs	495
Research Partners*, research	14 900
Enterprise partners**, research	1 529
Public partners***, research	2 799
Sum	**29 318**

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

PUBLICATIONS IN 2022

Journal and peer-reviewed conference papers

Aittokallio, T., 2022. What are the current challenges for machine learning in drug discovery and repurposing? *Expert opinion on drug discovery,* 17(5), pp.423-425.

Bakka, Haakon Christopher; Rognebakke, Hanne Therese Wist; Glad, Ingrid Kristine; Haff, Ingrid Hobæk; Vanem, Erik. Estimating the effect of biofouling on ship shaft power based on sensor measurements. Ship Technology Research (ISSN 0937-7255). 2022.

Bertinelli Salucci, Clara; Bakdi, Azzeddine; Glad, Ingrid Kristine; Vanem, Erik; De Bin, Riccardo. A novel semi-supervised learning approach for maritime lithium-ion battery monitoring. In: Proceedings of the 36th International Workshop on Statistical Modelling. EUT Edizioni Università di Trieste 2022 (ISBN 978-88-5511-309-0). pp 85-90. 2022.

Bertinelli Salucci, Clara; Bakdi, Azzeddine; Glad, Ingrid Kristine; Vanem, Erik; De Bin, Riccardo. Multivariable Fractional Polynomials for lithium-ion batteries degradation models under dynamic conditions. *Journal of Energy Storage* (ISSN 2352-152X). 2022.

Brandsæter, Andreas; Glad, Ingrid Kristine. Shapley values for cluster importance: How clusters of the training data affect a prediction. Data mining and knowledge discovery (ISSN 1384-5810).

Brant, S.B. and Hobæk Haff, I., 2022. The fraud loss for selecting the model complexity in fraud detection. *Journal of Applied Statistics*, pp.1-19.

Croella, Anna Livia; Dal Sasso, Veronica; Lamorgese, Leonardo Cameron; Mannino, Carlo; Ventura, Paolo. Disruption management in railway systems by safe place assignment. Transportation Science (ISSN 0041-1655) pp 938-952. 2022.

Dal Sasso, Veronica; Lamorgese, Leonardo; Mannino, Carlo; Tancredi, Antonio; Ventura, Paolo. Easy Cases of Deadlock Detection in Train Scheduling. Operations Research 2022 ; Volum 70.(4) s. 2101-2118

Eide, Simen; Leslie, David S.; Frigessi, Arnoldo. Dynamic slate recommendation with gated recurrent units and Thompson sampling. Data mining and knowledge discovery (ISSN 1384-5810). 36 pp 1756-1786. 2022.

Eliseussen, E., Fleischer, T. and Vitelli, V., 2022. Rank-based Bayesian variable selection for genome- wide transcriptomic analyses. *Statistics in Medicine, 41*(23), pp.4532-4553.

Engebretsen, S., Diz-Lois Palomares, A., Rø, G., Kristoffersen, A.B., Lindstrøm, J.C., Engø-Monsen, K., Kamineni, M., Hin Chan, L.Y., Dale, Ø., Midtbø, J.E. and Stenerud, K.L., 2023. A real-time regional model for COVID-19: Probabilistic situational awareness and forecasting. *PLOS Computational Biology,* 19(1), p.e1010860.

Hellton, K.H., Tveten, M., Stakkeland, M., Engebretsen, S., Haug, O. and Aldrin, M., 2022. Real-time prediction of propulsion motor overheating using machine learning. *Journal of Marine Engineering & Technology, 21*(6), pp.334-342.

Hjort, A., Pensar, J., Scheel, I. and Sommervoll, D.E., 2022. House price prediction with gradient boosted trees under different loss functions. *Journal of Property Research, 39*(4), pp.338-364.

Jalali, Neda; Brustad, Hilde Kjelgaard; Frigessi, Arnoldo; Macdonald, Emily Ann; Meijerink, Hinta; Feruglio, Siri Laura; Nygård, Karin Maria; Rø, Gunnar Øyvind Isaksson; Madslien, Elisabeth Henie; De Blasio, Birgitte Freiesleben. Increased household transmission and immune escape of the SARS-CoV-2 Omicron compared to Delta variants. *Nature Communications 2022;* Volum 13.(1) s. 1-5

Kloster, Oddvar; Mannino, Carlo; Riise, Atle; Schittekat, Patrick. Scheduling Vehicles with Spatial Conflicts. *Transportation Science 2022;* Volum 56.(5) s. 1165-1178

Lachmann, J., Storvik, G., Frommlet, F. and Hubin, A., 2022. A subsampling approach for Bayesian model selection. *International Journal of Approximate Reasoning, 151,* pp.33-63.

Lai, X., Taskén, H.A., Mo, T., Funke, S.W., Frigessi, A., Rognes, M.E. and Köhn- Luque, A., 2022. A scalable solver for a stochastic, hybrid cellular automaton model of personalized breast cancer therapy. *International Journal for Numerical Methods in Biomedical Engineering, 38*(1), p.e3542.

LeBlanc, M., Rueegg, C.S., Bekiroğlu, N., Esterhuizen, T.M., Fagerland, M.W., Falk, R.S., Frøslie, K.F., Graf, E., Heinze, G., Held, U. and Holst, R., 2022. Statistical advising: Professional development opportunities for the biostatistician. *Statistics in Medicine, 41*(5), p.847.

Lindstrøm, J.C., Engebretsen, S., Kristoffersen, A.B., Rø, G.Ø.I., Palomares, A.D.L., Engø-Monsen, K., Madslien, E.H., Forland, F., Nygård, K.M., Hagen, F. and Gantzel, G., 2022. Increased transmissibility of the alpha SARS-CoV-2 variant: evidence from contact tracing data in Oslo, January to February 2021. I*nfectious Diseases, 54*(1), pp.72-77.

Mannino, Carlo; Nakkerud, Andreas. Optimal Train Rescheduling in Oslo Central Station. Omega. *The* International Journal of Management Science 2022; Volum 116.

Olsen, Lars Henry Berge; Glad, Ingrid Kristine; Jullum, Martin; Aas, Kjersti. Using Shapley Values and Variational Autoencoders to Explain Predictive Models with Dependent Mixed Features. Journal of machine learning research (ISSN 1532-4435). 23(213) pp 1-51. 2022.

Parviero, Riccardo; Hellton, Kristoffer Herland; Haug, Ola; Engø-Monsen, Kenth; Rognebakke, Hanne Therese Wist; Canright, Geoffrey; Frigessi, Arnoldo; Scheel, Ida. An agent-based model with social interactions for scalable probabilistic prediction of performance of a new product. International Journal of Information Management Data Insights 2(2) 2022.

Pesonen, Henri; Simola, Umberto; Kohn Luque, Alvaro; Vuollekoski, Henri; Lai, Xiaoran; Frigessi, Arnoldo; Kaski, Samuel; Frazier, David T.; Maneesoonthorn, Worapree; Martin, Gael M.; Corander, Jukka. ABC of the future. International Statistical Review (ISSN 0306-7734). 2022.

Ravindran, V., Wagoner, J., Athanasiadis, P., Den Hartigh, A.B., Sidorova, J.M., Ianevski, A., Fink, S.L., Frigessi, A., White, J., Polyak, S.J. and Aittokallio, T., 2022. Discovery of host-directed modulators of virus infection by probing the SARS-CoV-2–host protein–protein interaction network. *Briefings in Bioinformatics, 23*(6), p.bbac456.

Salucci, C.B., Bakdi, A., Glad, I.K., Vanem, E. and De Bin, R., 2022. Multivariable Fractional Polynomials for lithium-ion batteries degradation models under dynamic conditions. *Journal of Energy Storage, 52*, p.104903.

Schenkel, Jonas F.; Zhang, Li-Chun. Adjusting misclassification using a second classifier with an external validation sample. Journal of the Royal Statistical Society Series A (Statistics in Society. 2022.

Thomas, Owen., Dutta, R., Corander, Jukka, Kaski, S. and Gutmann, M.U., Likelihood-free inference by ratio estimation. Bayesian Analysis, 17(1), pp.1-31. 2022.

Tjøstheim, D., Jullum, M. and Løland, A., 2023. Statistical Embedding: Beyond Principal Components. *Statistical Science, 1*(1), pp.1-29.

Tjøstheim, D., Jullum, M. and Løland, A., 2022. Some recent trends in embeddings of time series and dynamic networks. *Journal of Time Series Analysis*. 2022

Tveten, M., Eckley, I.A. and Fearnhead, P., 2022. Scalable change-point and anomaly detection in cross-correlated data with an application to condition monitoring. *The Annals of Applied Statistics, 16*(2), pp.721-743.

Vanem, Erik. Analyzing Extreme Sea State Conditions by Time-Series Simulation. I: *International Conference on Ocean, Offshore and Arctic Engineering ASME 2022 41st*. The American Society of Mechanical Engineers (ASME) 2022 ISBN 9780791885864

Vanem, Erik; Fazeres-Ferradosa, Tiago. A truncated, translated Weibull distribution for shallow water sea states. Coastal Engineering 2022; Volum 172.

Vitelli, Valeria; Fleischer, Thomas; Ankill, Jørgen; Arjas, Elja; Frigessi, Arnoldo; Kristensen, Vessela N.; Zucknick, Manuela. Transcriptomic pan-cancer analysis using rank-based Bayesian inference. Molecular Oncology (ISSN 1574-7891). 2022.

Zhang, Chi; Eken, Torsten; Jørgensen, Silje Bakken; Thoresen, Magne; Søvik, Signe. Effects of patient-level risk factors, departmental allocation and seasonality on intrahospital patient transfer patterns: network analysis applied on a Norwegian single-centre data set. BMJ Open (ISSN 2044-6055). 12(3) 2022.

Zhao, Z., Zucknick, M. and Aittokallio, T., 2022. EnrichIntersect: an R package for custom set enrichment analysis and interactive visualization of intersecting sets. Bioinformatics Advances, 2(1), p.vbac073.

Wahl, J.C., Aanes, F.L., Aas, K., Froyn, S. and Piacek, D., 2022. Spatial modelling of risk premiums for water damage insurance. Scandinavian Actuarial Journal, 2022(3), pp.216-233.

Reports and submitted papers

Bertinelli Salucci, C., Bakdi, A., Glad, I.K., Vanem, E. and De Bin, R., 2022. A novel semi-supervised learning approach for maritime lithium-ion battery monitoring. EUT Edizioni Università di Trieste

Boge Brant, S. and Hobæk Haff, I., 2022. Copulaboost: additive modeling with copula-based model components. arXiv e-prints, pp.arXiv-2208.

Hanes, R., Ayuda-Durán, P., Rønneberg, L., Zucknick, M. and Enserink, J.M., 2022. metascreen: A modular tool for the design and analysis of drug combination screens. bioRxiv, pp.2022-07.

Kamineni, M., Engø-Monsen, K., Midtbø, J.E., Forland, F., de Blasio, B.F., Frigessi, A. and Engebretsen, S., 2022. Effects of Non-Compulsory and Mandatory COVID-19 Interventions on Travel Distance and Time Away from Home: The Case of Norway in 2021. arXiv preprint arXiv:2206.03489.

Köhn-Luque, A., Myklebust, E.M., Tadele, D.S., Giliberto, M., Schmiester, L., Noory, J., Harivel, E., Arsenteva, P., Mumenthaler, S.M., Schjesvold, F. and Taskén, K., 2022. Phenotypic deconvolution in heterogeneous cancer cell populations using drug screening data. bioRxiv, pp.2022-01.

Liu, Q., Vitelli, V., Mannino, C., Frigessi, A. and Scheel, I., 2022. Pseudo-Mallows for Efficient Probabilistic Preference Learning. arXiv preprint arXiv:2205.13911.

Rønneberg, L., Kirk, P.D. and Zucknick, M., 2022. Dose-response prediction for in-vitro drug combination datasets: a probabilistic approach.

Storvik, Geir Olve; Palomares, Alfonso Diz-Louis; Engebretsen, Solveig; Rø, Gunnar Øyvind Isaksson; Engø-Monsen, Kenth; Kristoffersen, AB; De Blasio, Birgitte Freiesleben; Frigessi, Arnoldo. A sequential Monte Carlo approach to estimate a time varying reproduction number in infectious disease models: the Covid-19 case. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2022 (to appear).

Wahl, Jens Christian; Aas, Kjersti. Modeling claim frequency for car insurance using meteorological data. Norsk Regnesentral, NR-notat SAMBA/09/22. pp 36. 2022.

Wahl, Jens Christian; Aas, Kjersti. Modeling claim frequency for car insurance using meteorological data. Norsk Regnesentral, NR-notat SAMBA/22/22. pp 36. 2022.

Open source published software

BayesMallows: Bayesian Preference Learning with the Mallows Rank Model [CRAN] [GitHub] [R Journal]

BayesSUR: Bayesian Seemingly Unrelated Regression [CRAN]

bayesynergy: An R package for Bayesian semi-parametric modelling of in-vitro drug combination experiments [GitHub]

DIscBIO: A user-friendly R pipeline for biomarker discovery in single-cell transcriptomics [GitHub]

ELFI: A statistical software package for likelihood-free inference [Zenodo] [Read the Docs] [GitHub]

EnrichIntersect: an R package for custom set enrichment analysis and interactive visualization of intersecting sets [CRAN] [GitHub]

hdme: High-Dimensional Regression with Measurement Error [CRAN] [GitHub] [Journal of Open Source Software]

kdensity: An R package for kernel density estimation with parametric starts and asymmetric [CRAN] [GitHub] [Journal of Open Source Software]

matlab2r: Translation Layer from MATLAB to R [CRAN] [GitHub]

MCCE: Monte Carlo sampling of realistic Counterfactual Explanations for tabular data
[GitHub R] [GitHub Python]

pycox: Survival analysis with PyTorch [GitHub] [PyPI]

pyPhenoPop: Phenotypic deconvolution in heterogeneous cancer cell populations using drug screen data
[PyPI] [GitHub]

shapr: Explaining the output of machine learning models with more accurately estimated Shapley values [CRAN] [GitHub] [Journal of Open Source Software]

Shiny app for prior elicitation [shinyapps.io] [GitHub]

spread: An R package that contains different infectious disease spread models [CRAN] [GitHub]

streamchange: A package for segmenting streaming time series data into homogenous segments [GitHub]

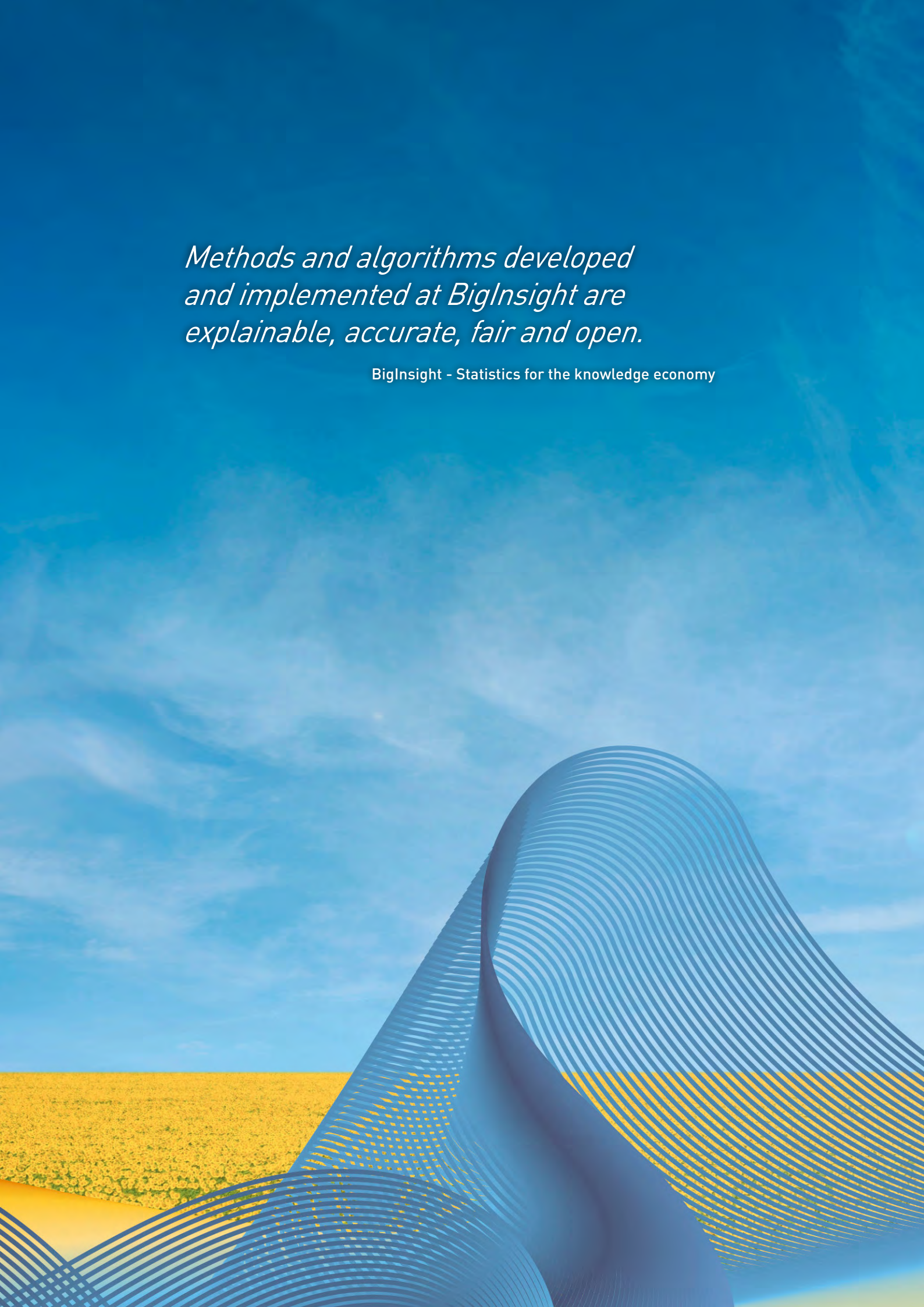
tpca: automatically selecting the principal components most sensitive to changes [GitHub]

tpcaMonitoring: performing TPCA change detection
[GitHub]

twl: Two-Way Latent Structure Clustering Model [CRAN]

Methods and algorithms developed and implemented at BigInsight are explainable, accurate, fair and open.

BigInsight - Statistics for the knowledge economy



Biginsight.no

Postal address

PO box 114 Blindern
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

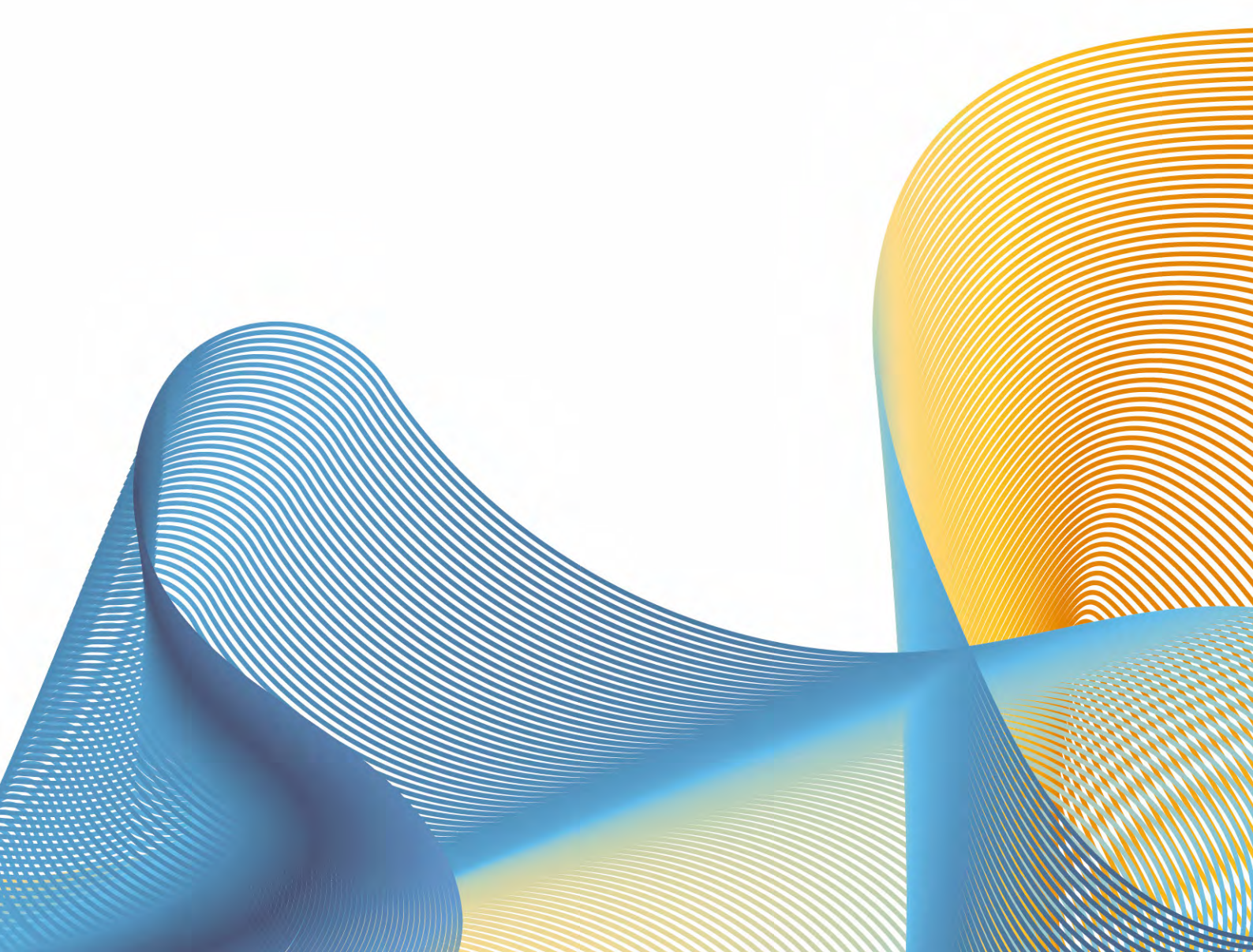
Arnoldo Frigessi +47 95735574
Norsk Regnesentral +47 22852500

Photocredits:
Page 1 and 66, 6, 16-17, 27, 36-37, 62-63: Shutterstock. Licensed under Creative Commons / norden.no: Page 29: Nikolaj Bock - norden.org, Page 30-31: Nikolaj Bock – norden.org, page 50-51 Karin Beate Nøsterud – norden.org. All other photos if not listed here or under photo: BigInsight

Design og lay-out:
Designavdelingen / Kunst i Skolen. Gjennom å bruke Kunst i Skolens designavdeling støtter BigInsight Kunst i Skolens arbeid for kunstformidling og kreativitetstrening for barn og unge. kunstiskolen.no

In 2022 a terrible war is again burning in Europe. We stand with the Ukrainian people, who are defending their country and their freedom, and hope that peace will win again.

BigInsight's annual report 2022 is coloured throughout with the colours of Ukraine.




BigInsight