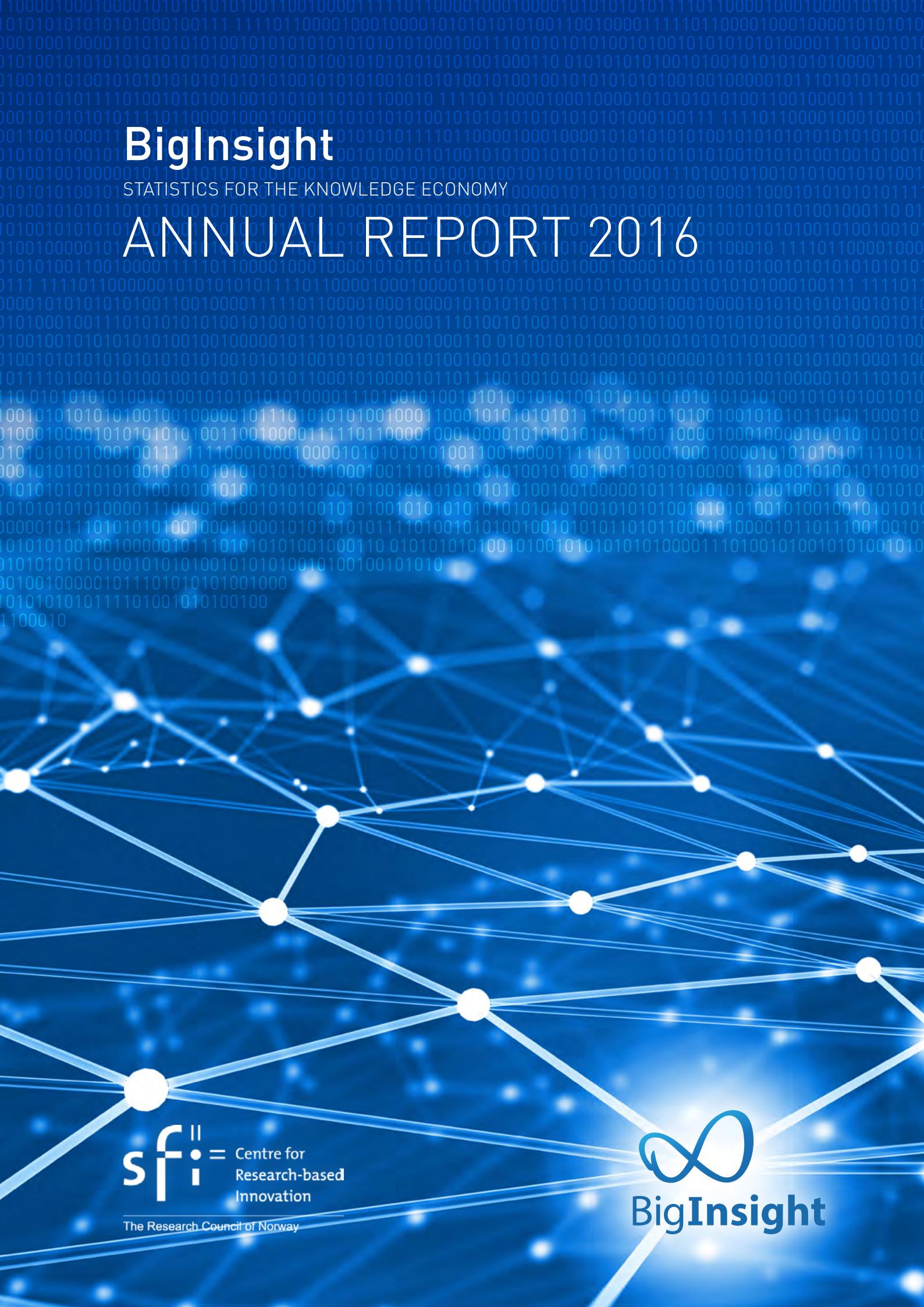
BigInsight

STATISTICS FOR THE KNOWLEDGE ECONOMY

ANNUAL REPORT 2016



sfi Centre for Research-based Innovation

The Research Council of Norway

BigInsight

Statistics is the science of learning from data, and accounting for relevant uncertainties. As such, it permeates the physical, natural, and social sciences, as well as public health, medicine, technology, business, and policy

(American Statistical Association)

CONTENT

SUMMARY

BigInsight produces innovative solutions for key challenges facing a consortium of private and public partners, by developing and applying original statistical and machine learning methodologies. Exploiting huge and unique data resources and substantial scientific, industrial and business knowledge, we construct personalised solutions, predict dynamic behaviours and control processes that are the core of the partners' innovation. We shall discover radically new ways to target products, services, prices, therapies, technologies, towards individual needs and conditions, thus providing improved quality, precision 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, etc. This is possible thanks to the unprecedented availability of large scale measurements and individual information together with new statistical theory, computational methods and algorithms able to extract knowledge from complex and high dimensional data. When we develop methods and algorithms we consider five principles: responsibility, explainability, accuracy, auditability and fairness.

Research at BigInsight will lead to value creation and will help bring our partners to new leading positions. In the era of digitalization, BigInsight creates unique competence and capacity for the Norwegian knowledge-based economy, contributing to the development of a sustainable and better society.

BigInsight started in April 2015 as one of the third generation Norwegian Centres for Research-based Innovation. It is funded by the Research Council of Norway and by fourteen partners and will operate until 2023. This is the annual report of the second year of the center.

[The development of theory, methodologies and algorithms] will inform scientific and technological discoveries, create new business opportunities, accelerate solutions to global challenges, inform policymaking, and improve the environment, health and infrastructure of the world in an 'Age of Algorithms'

(The Turing Institute, UK)

VISION AND OBJECTIVES

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

There is a dramatic scope for industries, companies and nations – including Norway – to create value from employing novel ways of analysing complex data. The complexity, diversity and dimensionality of the data, and our partner's innovation objectives, pose fundamentally new challenges to statistical inference. We develop original, cutting-edge statistical, mathematical and machine learning methods, produce high-quality algorithms implementing these approaches and thereby deliver new, powerful, and operational solutions.

BigInsight's research converges on two central innovation themes:

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

Our solutions are significantly better than the state-of-the-art, thanks to brilliant, courageous and creative generic methodologies that extract knowledge from complex data. Generic methodology and their new applications are published on 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.

In the next 50 years, ample data will be available to measure the performance of algorithms across a whole ensemble of situations. This is a game changer for statistical methodology. Instead of deriving optimal procedures under idealized assumptions within mathematical models, we will rigorously measure performance by empirical methods, based on the entire scientific literature or relevant subsets of it.

David Donoho, 50 Years of Data Science. (2015)

Personalised solutions

The core operation of our partners involves interacting with many individual units: at Telenor, millions of individual mobile phone customers are part of a communication network; at Gjensidige, a million policyholders share risks of contingent, uncertain losses; for DNB, customers transfer money and receive loans; at OUS, cancer patients need to be treated in the most effective personalized way; the Norwegian Insititute of Public Health works with individuals susceptible for infections or infectious themselves; NAV supports hundreds of thousands of people with special needs in relation to health or employment in challenging life situations; for Skatteetaten, millions of taxpayers need to be mentored to declare their incomes; for DNV GL and ABB, hundreds of sensors register the functional state and operation of a ship at sea; for DNV GL and OUS, a multitude of sensors monitor safety in healthcare as patients' health records are digitalised. There are many common characteristics:

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

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

Predicting transient phenomena

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

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

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

Photo: Igor Stevanovic / Shutterstock.com

ORGANISATION

Board in 2016

Tron Even Skyberg, DNB, chairman from 25.11.16
Tone S. Eilertsen, Gjensidige, chairman until 25.11.16
Rune Braastad, ABB
Bobbie Nicole Ray-Sannerud, DNV GL
Birgitte F. De Blasio, Folkehelseinstituttet
Erlend Willand-Evensen, Gjensidige, from 25.11.16
Stefan Erath, Hydro
Ludvig Guldal, NAV
Lars Holden, Norsk Regnesentral
André Teigland, Norsk Regnesentral
Peder Heyerdahl Utne, Oslo University Hospital
Marcus Zackrisson, Skatteetaten
Astrid Undheim, Telenor
Bård Støve, University of Bergen
Arne Bang Huseby, University of Oslo

Observer: Terje Strand, Research Council of Norway
The board had 2 meetings in 2016. All partners are represented in the Board.

Legal organisation

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

Director

Arnoldo Frigessi, UiO
UiO Director

Co-Directors

Ingrid Glad, UiO, Co-Director
Kjersti Aas, NR, Co-Director
André Teigland, NR,Co-Director
Håvard Rue, NTNU, until 31.12.16
Anders Løland, NR Co-Director, from 01.01.17
Ingrid Hobæk Haff, UiO, Co-Director, from 01.01.17

Principal Investigators

Magne Aldrin, NR
Ingrid Hobæk Haff, UiO
Alex Lenkoski, NR
Arne Bang Huseby, UiO, until 31.12.16
Anders Løland, NR
Arnoldo Frigessi, UiO
Carlo Mannino, UiO, from 01.01.17
Ingrid Glad, UiO Håvard Rue, NTNU, until 31.12.16
Clara Cecilie Günther, NR
Magne Thoresen, UiO

Administrative Coordinator

Unni Adele Raste, NR

Scientific Advisory Committee (SAC)

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

BigInsight

BOARD

NR´S RESPONSIBLE PERSON — ADVISORY SCIETIFIC COMMITEE

DIRECTOR

CO-DIRECTORS

ETHICAL ADVISING GROUP

PRINCIPAL INVESTIGATORS AND CO-PRINCIPAL INVESTIGATORS

INNOVATION OBJECTIVES

Personalised marketing

Personalised health and patient safety

Personalised fraud detection

Sensor systems

Forecasting power systems

PARTNERS

Partners

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

Cooperation between the partners of BigInsight

In addition to the board meetings, where all partners are represented, we organized several thematic seminars across the Innovation Objectives in 2016. These seminars focused on network analysis, anomaly detection and analysis of web clicking data. Within the Innovation Objectives there are frequent joint working sessions and regular update and progress meetings where the partners play a central role.

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

RESEARCH STRATEGY

Within our two innovation themes, we have identified five innovation objectives to harbour all research at BigInsight. Each innovation objectives (IOs) is centered on a concrete innovation area: marketing, health, fraud, sensor, power. Most partners join more than one IO (and changes will happen in the years to come, as focus and interests will develop further). Each IO has a few specific innovation aims related to outstanding unresolved problems, which we believe can specifically be solved with new statistical, mathematical and machine learning methodologies. Our research projects are designed to deliver methods and tools for their solution. We aim to new, interesting and surprising solutions, which take the field and our partners ahead in their innovation agenda. Final transfer to partners' operations will happen both within and on the side of BigInsight.

INNOVATION OBJECTIVES

Personalised marketing

Personalised health and patient safety

Personalised fraud detection

Sensor systems

Forecasting power systems

INNOVATION PARTNERS

DNB	DNV-GL	DNB	ABB	DNV-GL
Gjensidige	Kreftregisteret	Gjensidige	DNV-GL	Hydro Energy
Skatteetaten	OUS	NAV		
Telenor	Telenor	Skatteetaten		

RESEARCH PARTNERS

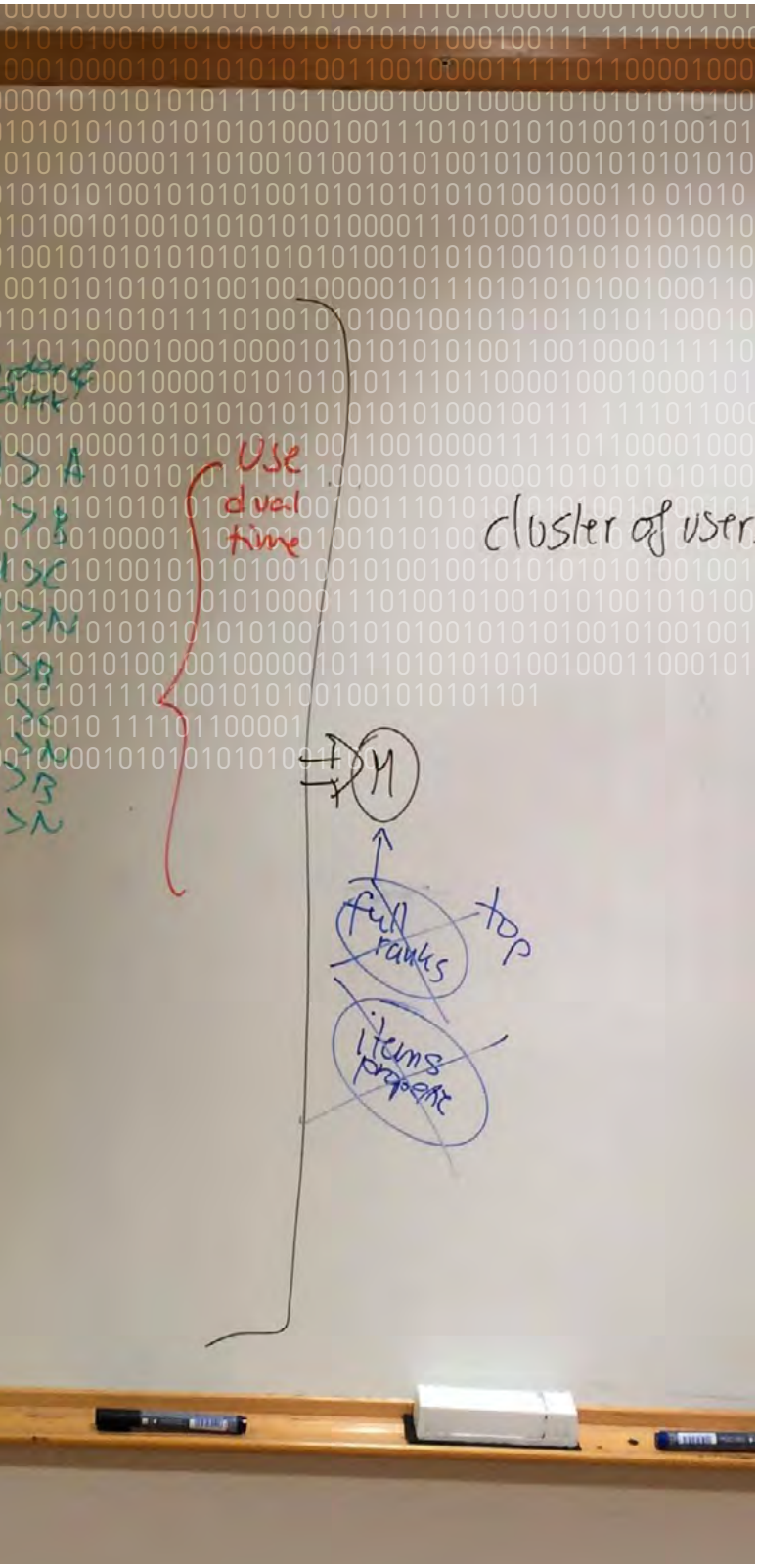
NR	UiO	NR	NR	NR
UiO	OUS	UiO	UiO	UiO
NIPH	NR	UiB	UiB	UiB
UiB	UiB			
	NIPH			

PRINCIPAL INVESTIGATORS

Principal Investigators:	Kjersti Aas	Magne Thoresen	Anders Løland	Ingrid Glad	Alex Lenkoski
co-Principal Investigators:	Arnoldo Frigessi	Clara Cecilie Günther	Ingrid Hobæk Haff	Magne Aldrin	Carlo Mannino

METHODS

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



Sampling bias and missing values take new dimensions

Data can be collected for one first purpose (say, billing) and then used in a different context (marketing): bias must be corrected. Data composed by several collections or collected in different periods can be inconsistent, something which has to be resolved. The unsolicited production of data on a volunteer basis can be utilised only if properly calibrated. More generally, data are often fragmented, with individuals entering and leaving cohorts and surveys at different time points and for different reasons, generating informative missingness. An arduous situation is encountered when data are only temporarily available, and are then deleted (because of regulations), or arrive in data streams, requiring the collection of appropriate statistics for future use.

There are new possibilities in using all data

Traditionally, data quality is strongly advocated, so that "bad" data might even be best omitted. For data rich areas this approach is too conservative, and we will design stochastic models to explore the information hidden in the complete mosaic of data. By integrating more data layers, we will be able to make more precise inference, combination of many weak signals and complex interactions. Bayesian borrowing of strength can compensate for data quality or uneven quantity. It also allows the rigorous integration of data with substantive knowledge, both hard facts as constraints and soft expertise as elicited prior models.

High frequency time series data allow intervention in real time

Sensors, machines, patients and customers generate very high dimensional time series which are analysed for motif discovery, anomaly detection and classification. The purpose is to automatically alert about shifts in trends, variability or extremes, about changing patterns of behaviour or about any potential deviations from the norm. Most time series capture the system in a transient phase and are therefore not stationary and with strong dependences. Sensor data, in industry and healthcare, have a time resolution which can be adaptively increased when anomalies appear as forthcoming or reduced in normal periods. In order to reduce false alarms, methods may effectively evoke sparsity assumptions. Measuring reactions to interventions, allows the design of new experimental plans, where groups of individuals are exposed to new situations (products, therapies), and observed in their transient phase.

Dimension reduction
Both the dimension of the data and of the unknown parameters space can be huge. In the first case, one might need to subsample the data or to distribute them to computers working in parallel, asynchronously. In the second case, one need to determine a reduced model which still allows efficient prediction. Penalised likelihood based approaches, or based on approximate models, but also boosting and other machine learning approaches are very useful.

Causal effects enable effective actions
Causal statistical modelling is opening new avenues in data analysis, moving from pointing to association to estimating causes. Graphical, counterfactual and networks models allow to ask what would have happened if an intervention had not been made? While counterfactual cases cannot be observed, some knowledge about them can usually be inferred from the data collected. Time-dependent confounding is a particular challenge. The distinction between direct and indirect effects allows interpreting causes.

Rapid computation means model approximation
Real time computations permit real time interventions. This is difficult, when large quantities of data need to be analysed or when the space of solutions is huge. Trading model accuracy and inferential precision for efficiency becomes imperative. Parallel computing, asynchronous and synchronous, on varying architectures, open new possibilities for big data analysis. We experiment with various solutions and are particularly interested in subdividing covariates (not samples) between machines in a cluster. Pseudolikelihood approximations should play a new role as approximate models, as they are efficiently estimated. Approximate Bayesian Computation and its variations might scale with problem dimension.

Anomaly detection and changepoint prediction allows control
Measures of surprise quantify the level of incompatibility of data with a given model, without any reference to alternatives. Surprise plays an important role in dynamic situations, where the reference is the past trajectory. There are connections to outlier theory; measures of Bayesian discrepancies between priors and posteriors; change point detection and we will need to extend and adapt these ideas to highly multivariate and non-stationary time series. A further aim is to develop models that allow change point prediction, rather than merely locating these after occurrence. We shall develop procedures to monitor parallel streams of data to detect anomalies and change-points.

Network based decision theory
Stochastic and dynamic networks appear in many innovation objectives, linking units by similarity, proximity or contact. The study of the mechanisms governing network growth is important and allows prediction and personalised intervention. We wish to develop a decision theory that exploits network structure, whenever decisions are taken according to local latent communities.

Large Scale Optimisation
Discrete and continuous optimisations are central tasks in decision making, management and inference. Highly efficient in the linear and convex case, optimisation becomes very hard in non-convex situations and in high dimensions. For example, many network based problems are large-scaled, NP-complete combinatorial optimisation problems which can only be tackled by suitable decomposition methods or efficient approximate algorithms. By using a mix of exact decomposition methods and heuristics, large optimization problems previously regarded as intractable are however now feasible.

Deep Learning
Deep learning allows precision classification in supervised problems, without the need to careful modelling. It allows optimising feature selection including interactions and non-linear effects. However, it requires very large amounts of labelled training data and computer power to train. We explore the use of deep learning in various situations, investigate power and scaling properties.

SCIENTIFIC ACTIVITIES

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

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

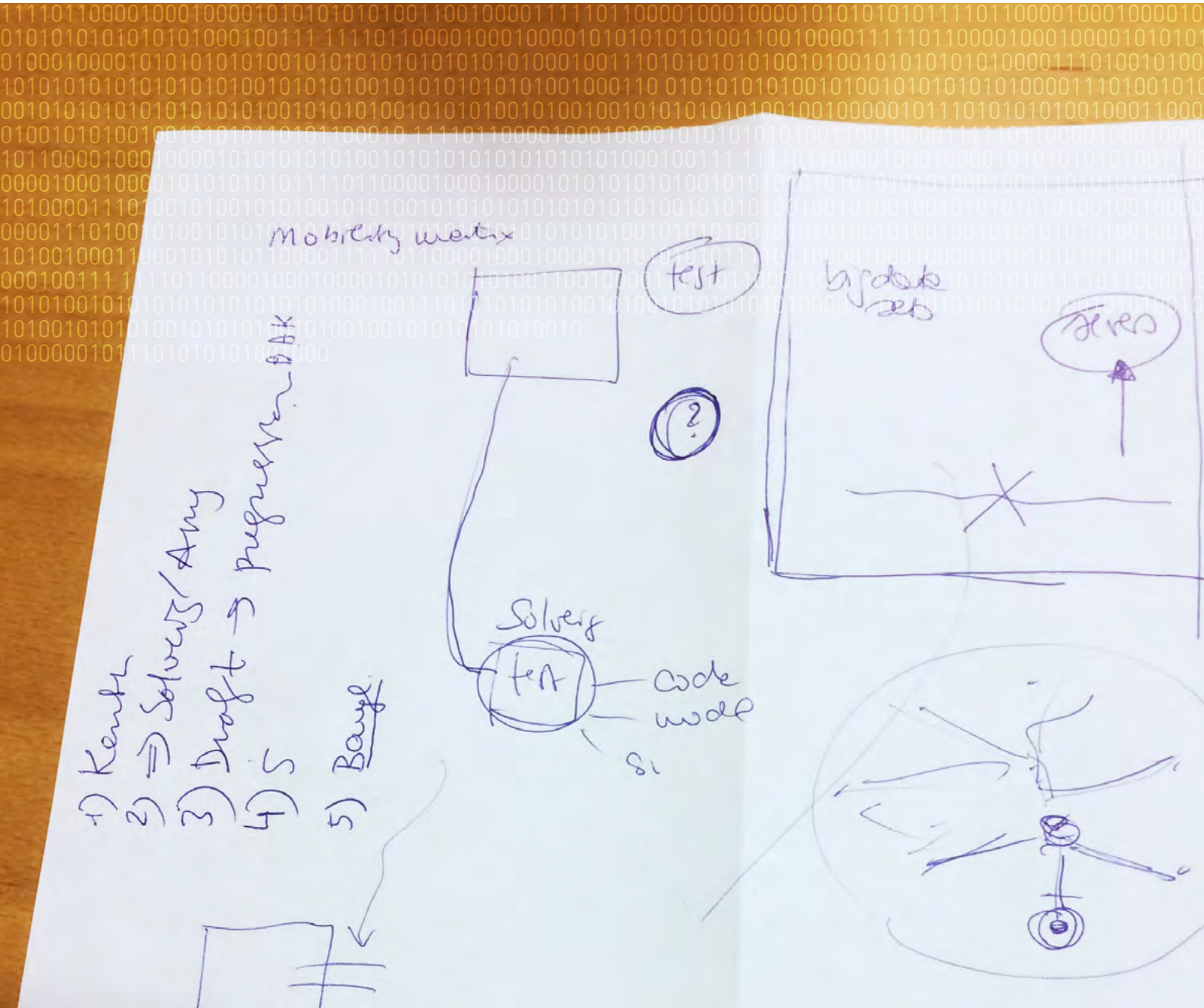

Mobility matrix
test
Solvers
ten
code
mod
1) Kandi
2) ⇒ Solvers/Array
3) Draft ⇒ preserve
4) S
5) Bayes

Photo: Shutterstock

PERSONALISED MARKETING

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

Achievements 2016

Stochastic customer growth dynamics

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

Predicting customer behavior from time series

Multiple time series, related to individual customers have been used to predict, as early as possible, whether the customer will pay back his loan or not. We have developed a new solution based on deep learning of time series, which gives excellent results on the DNB cases it has been tested on. The method will be transferred to DNB and published.

Bayesian methodology for recommender systems

In many important situations, we wish to recommend to each individual user or customer, the items she might be most interested in, or the ones she would benefit most. Starting from data where user either rate or compare items or click on them, we want to predict user's preferences on other items. In some applications, we are also interested in estimating the shared consensus preference of a homogeneous group of people. We have invented a new Bayesian approach based on extensions of the Mallows model, which allow making individualized recommendations, which are equipped with a level of uncertainty. We are working on public data, NRK data and plan to analyse a case from Telenor. The methodology was nominated to the Inven2 Idéprisen 2016 for the three best innovation projects at the University of Oslo and the Oslo University Hospital.

Stochastic models for early prediction of viral customer behavior on networks

Can we predict if a new digital service or product will be widely adopted in a market, or if it will be a flop? We look to markets where adopters are organised as nodes in a network, where links represent contacts which allow a user to "convince" a neighbor to adopt. We work on a case from Telenor. The aim is to be able to predict, as early as possible, how the adoption process will spread virally (or not) on the network. Our Bayesian simulation based approach looks very promising.

Activities 2017

- Bayesian methodology for recommender systems
- Stochastic models for early prediction of viral customer behavior on networks
- Network dynamics
- Analysis of clickstream data in order to identify customer segments
- Default prediction using network data on companies

Principal Investigator
Kjersti Aas

co-Principal Investigator
Arnoldo Frigessi

PERSONALISED HEALTH AND PATIENT SAFETY

The health system is producing data at an unre- strainable 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 prototype projects.

Standard cancer treatment consists of a combination of drugs, at various dosages and in different order, implying a huge number of possible cocktails. We develop a new approach to therapy optimization, based on simulation of cancer growth. Also, we develop methods to predict the synergy between drugs based on cancer cell line data. This allows exploring new approaches to personalized cancer treatments.

National population based cancer registries routinely publish survival statistics. To become more relevant both to the clinician and the patient, the survival statistics should preferably be tailored to encompass more detailed information, moving towards personalized cancer statistics.

Patient safety is critical in healthcare. The amount of data collected in healthcare is vast and rapidly expanding, including electronic health records (EHR) and health care system/organization data. We develop new data analytic methods to predict and control risk in healthcare organizations at system level.

We use data on mobile phone locations and their movements to describe movement of and contacts between people, with the purpose to monitor and predict the development of infectious diseases. Simulating various vaccine plans allows evaluating cost-benefits.

Personalized cancer statistics
National population based cancer registries publish survival statistics by cancer site, stage, gender and time period, using relative survival methods. As new clinical registries are established, more data on treatment and later events become available, in addition to information on comorbidity or income and educational level. To become more relevant both to the clinician and the patient, the survival statistics will be tailored to encompass more detailed information, in line with the tradition of risk prediction models.

Personalized cancer treatments
We develop a new mathematical, statistical and computational methodology to improve current ways of predicting treatment response for breast cancer patients. We design theoretical and computational models that capture key molecular and cellular mechanisms of the breast cancer, reproducing the effect of specific treatments used in actual clinical trials. The models are personalized to specific subgroups by calibrating the computer simulations to actual patient data. This allows personalized treatment guidance, by simulating an array of possible treatment schedules for a given patient to find the optimal regime. In addition, we develop and extend statistical learning methods for predictive modelling of cancer drug sensitivity based on large-scale in vitro screens of drugs and drug combinations.

Healthcare safety management
This project harvest information lying in the collections of high dimensional health records and administrative databases, routinely acquired in hospitals and health institutions, which carry a preventive signal relative to a potential harm to patients. This signal is used to predict the occurrence of a possible harm in an automatic way at aggregated health institution/ward level. It can also be used to define new control rules on hazards, to prevent and mitigate the risk of harm.

Telecom data for epidemics control
We study the spread of infectious diseases, by observing social mixing and mobility patterns of susceptible and infectious individuals, who are key drivers of the spatial dissemination of infectious diseases. The use of mobile phone data containing geo-temporal information from individuals gives an accurate, real-time description of population movements. This in turn allows accurate predictions of epidemic spread.

Activities 2017

- Variable selection in highly dimensional time series regressions
- Multi-state Markov models with frailties
- Stochastic model of breast cancer growth
- First pilot study of hazard/harm prediction at system level for a hospital ward
- Mobile movement data in Norway for infectious diseases prediction
- Drug synergy prediction in cancer cell lines by data integration techniques.

Principal Investigator
Magne Thoresen

co-Principal Investigator
Clara Cecilie Günther

Photo: Shutterstock

PERSONALISED FRAUD DETECTION

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

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

Ensemble methods for fraud detection
Fraud detection can be seen as a regression/forecasting problem, where fraud (true/false) is the response, possibly with a potential economic loss, and there are very many covariates. Including interactions, the number of covariates is huge. Generally, there are few fraud cases that are investigated, and a great number of undetected cases exist. The objective is to produce a trustworthy probability of fraud for each case. Many statistical and machine learning methods already exist. Combining results produces better results. We construct a toolbox for combining fraud forecasting models, exploiting both the time series aspect of the data and the covariates, in addition to the probabilities stemming from each individual model.

Text-mining for fraud detection
In addition to ordinary variables (age, demography, background, behaviour, etc.), a potential fraud case can be accompanied with a variable amount of text, for example the policy holder's description of a claim or an officer's summary of a case. These texts are informative for a human eye searching for fraud. The objective is to exploit recent advances in text-mining to produce text related features that can be used in the statistical models, and to investigate the added value of these.

Network analysis for fraud detection
Fraud can be viral, spreading directly or indirectly from one fraudster to others. Exploiting knowledge about social relations can be useful. Understanding how such networks of users look and evolve over time is expected to significantly improve fraud detection models. We build these networks and extract useful characteristics to produce better fraud forecasts and provide additional insight into how fraud spreads.

Achievements 2016

- Text mining seminar held, but text mining analysis postponed due to lack of data
- Prototype toolbox for fraud detection
- Initial misreporting of VAT and prepared risk for future misreporting of VAT project
- Prototype detection of money laundering and detection of insurance fraud
- Continuous work on research applications for access to various data

Activities 2017

- Seminar on network analysis for fraud detection
- Methods and tools for variable selection
- Exploit network relations between individuals and businesses in statistical models
- New, local discriminant methods
- Fraud detection toolbox improved, with both ensemble and network models
- Misreporting of VAT and risk for future misreporting of VAT
- Detection of money laundering, insurance fraud and social security fraud

Principal Investigator
Anders Løland

co-Principal Investigator
Ingrid Hobæk Haff

SENSOR SYSTEMS

Maintenance and inspections of ships are traditionally based on a preventive scheme where components have been maintained according to a time schedule. This approach is based on the assumption that a component has a defined lifetime, after which its failure rate increases. However, estimates of lifetime have large uncertainties and a large percentage of failures are not age-related and are therefore not adequately addressed by preventive scheduled maintenance. We develop new approaches based on the recent availability of large arrays of sensors, which monitor condition and operation of machinery and power systems. Sensor data are becoming available for the first time on global ship fleets, with efficient broadband connectivity to shore. We suggest new approaches to condition 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. For the design of sensor monitoring systems, a key challenge is to determine the level of resolution in time and sensor density needed to have a precise dynamic picture of the actual health of the system. Borrowing strength across sensors and ships in a fleet is an important aspect, leading to increased safety of a whole fleet.

Anomaly detection via signal reconstruction based on the past

We develop a multi-sensor, multi-scale statistical approach to detect, as rapidly as possible, anomalies. Data originate from sensors covering different aspects of a ship operation (f.ex. propulsion power, speed over ground and ship motion in four degrees of freedom). The developed method employs Auto Associative Kernel Regression (AAKR) for signal reconstruction, and the Sequential Probability Ratio Test technique for anomaly detection, where different hypothesis tests looking both at mean and variance deviations have been tested. We demonstrate that our model produces good reconstructions and as long as the parameters are tuned carefully, alarms are triggered appropriately.

Models for ship operations and efficiency

Understanding ship-internal and external factors that regulate ship propulsion and fuel efficiency is important to optimise ship design and operations. Deviations between the measured ship speed and the ship speed predicted on the basis of propulsion power and other internal and external factors, is an indication of a possible anomaly, incl. hull, propeller or engine damage. We compared statistical models which allow such prediction, using also external environmental factors that clearly affect sensor data. We are able to produce very good predictions, which can be further used for anomaly detection. A similar approach was also successfully tested on an cooling system for a motor.

Activities in 2017

- Unsupervised anomaly detection in sequential categorical log data
- Surprise detection for sub systems by multivariate multiple regression
- Dimension reduction procedures for changepoint detection
- Improving the AAKR test component
- Hull condition monitoring, sensor based and virtual

Principal Investigator
Ingrid Glad

co-Principal Investigator
Magne Aldrin

FORECASTING POWER SYSTEMS

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

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

Error dressing spot price forecasts

Electricity markets "spike" and "crash" when volumes are respectively slightly higher or lower than typical and these extreme price swings make uncertainty quantification a critical part of forecasts. However, the limited degree to which these extremes are observed makes such constructions difficult. We use published bid/ask curves that determine the final price, to construct realistic distributional price forecasts that embed this extreme behaviour. We employ the concept of "error dressing" by using the curves to translate residual behaviour of market volume forecasts into price uncertainties.

Using river inflow projections to augment electricity price spot forecasts

River inflow is a critical quantity that impacts the bidding strategies of hydroelectric operators. We use newly available data on river inflows for Norway and Sweden to build a river inflow component into spot price forecasting system.

Power Matching Problem for the Smart Grid

The Power Matching Problem amounts to determine the optimal (according to user preferences) e-profile, that is the utilization of the available energy during the time horizon by the appliances in the households so that demand and supply meet and all time constraints are satisfied. As such, this can be seen as a Resource Constrained Scheduling Problem with Time Windows and in presence of uncertainty. We can use stochastic optimization, where one minimizes the expected value of a given objective function, based on some probability distribution attached to a set of possible input scenarios. Alternatively, robust optimization does not need such a probability distribution, as just ranges of operations are needed.

Transfer of methodology to Hydro

In 2016 we implemented an "error dressing" methodology that will serve as the core of our distributional forecasting system. This was first implemented for the Nordic electricity market and shared with the industrial partner, Hydro. The results were so promising that Hydro has already put the methodology into production and multiple individuals at Hydro use the model output to inform production and trading decisions on a daily basis.

Activities 2017

- Expansion of the error dressing methodology to all other relevant European electricity markets, in particular the German (EEX) market.
- River inflow model for the Nordpool spot market.
- Electricity flow model to capture trading behavior between markets
- Curve disaggregation model to take published system-level bid/ask data and construct regional bid/ask curves
- Power Matching Problem for the Smart Grid

Principal Investigator
Alex Lenkoski

co-Principal Investigator
Carlo Mannin

INTERNATIONAL COOPERATION

International Academic Partners

International Academic Partners contribute to place BigInsight in the center of the global data science community. We collaborate in research and co-supervise PhD students. Joint workshops and training events will be organised.

STOR-i, Statistics and Operational Research in partnership with Industry, is a joint venture between the Departments of Mathematics & Statistics and Management Science of the University of Lancaster. STOR-i offers a unique interdisciplinary PhD programme developed and delivered with important UK industrial partners. The centre is at the forefront of international research effort in statistics and operation research, establishing an enviable track record of theoretical innovation arising from real world challenges. Professors Jonathan Tawn, professor Idris Eckley (who co-lead the centre) and professor Paul Fearnhead co-supervise PhD students together with BigInsight staff, on particle filters, multivariate extremes, non-parametric isotonic spatial regression, Bayesian modelling, multivariate sensor data. BigInsight and STOR-i exchange membership in each other's scientific advisory boards.

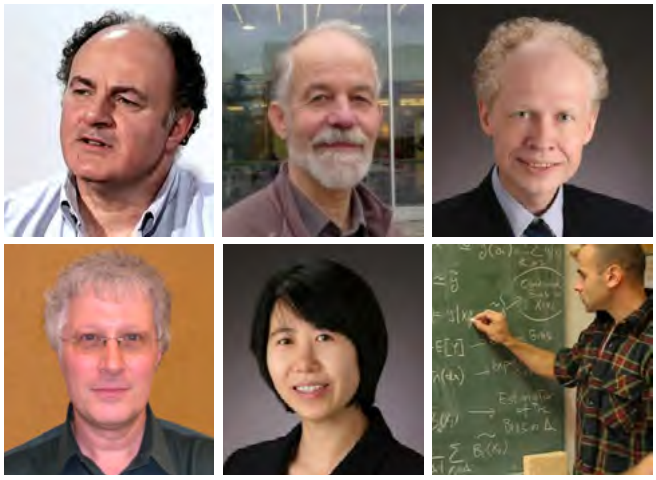
Professors Idris Eckley, Jonathan Tawn and Kevin Glazebrook, leading STOR-i at University of Lancaster

Professor Sylvia Richardson, MRC Biostatistics Unit, Cambridge

The Medical Research Council Biostatistics Unit (BSU) is part of the University of Cambridge, School of Clinical Medicine. It is a major centre for research, training and knowledge transfer, with a mission 'to advance biomedical science and human health through the development, application and dissemination of statistical methods'. BSU's critical mass of methodological, applied and computational expertise provides a unique environment of cutting edge biostatistics, striking a balance between statistical innovation, dissemination of methodology and engagement with biomedical and public health priorities. Professor Sylvia Richardson is director of the BSU. BigInsight and the BSU have several joint projects in health and molecular biology. Professor Richardson will receive an honorary doctorate of the University of Oslo in 2017.

International guest programme

BigInsight has an international guest programme, which includes all from short visits to long term visiting and adjunct positions and a sabbatical visitor programme.
In 2016 we hosted the following longer visits:
Professor Gianpaolo Scalia Tomba, University of Roma Tor Vergata, visits Oslo regularly in collaboration with NIPH (Health)
Professor Emeritus Elja Arjas, University of Helsinki, has a 20% adjunct position at BigInsight.
Professor Per Mykland, University of Chicago, spends a sabbatical year (2016-2017) at BigInsight (Sensor)
Professor Peter Müller, University of Texas, spent the autumn semester 2016 at BigInsight (Fraud, health)
Professor Lan Zhang, University of Illinois at Chicago, spends a sabbatical year at BigInsight (Sensor)
Associate Professor Fabio Divino, University of Molise, Italy (Health)



Professor Gianpaolo Scalia Tomba,
Professor Emeritus Elja Arjas,
Professor Per Mykland,
Professor Peter Müller,
Professor Lan Zhang,
Associate Professor Fabio Divino

International training programme

PhD students from other universities can spend periods of training and research collaboration at BigInsight.
In 2016 we welcomed
Marta Crispino, University Bocconi, Milano (Recommendation systems)
Denekew Belay Bitew, Hawassa University (Malaria models and world mortality rates)
Derbachew Asfaw, Hawassa University (Dynamic rank data)
Dorothea Weber, Berlin (Drug effect prediction)

BigInsight is a partner of the Norwegian Programme for Capacity Development in Higher Education and Research for Development (NORHED) project at Hawassa University (Ethiopia), together with NTNU.
https://www.norad.no/en/front/funding/norhed/pro- jects/hawassa-university--phd-programme-in-math- emati-cal-and-statistical-sciences/

International Programmes and Funding

BigInsight is partner of the COST Action CA15109 "European Cooperation for Statistics of Network Data Science (COSTNET)". Professor Arnoldo Frigessi is nominated as a Member in the Management Committee and professor Birgitte Freiesleben de Blasio is nominated as deputy. This EU action started in 2016 and aims to facilitate interaction between diverse groups of statistical network modellers, establishing a large and vibrant interconnected and inclusive community of network scientists. The first workshop took place in Slovenia.
http://www.cost.eu/COST_Actions/ca/CA15109

cost
EUROPEAN COOPERATION IN SCIENCE AND TECHNOLOGY

Network of Big Data Centers of Excellence in Europe

Big Data National Centers of Excellence in Europe join forces for better networking and collaboration. BigInsight is part of the consortium. The main focus of the network is research itself and how research can be transferred into relevant industries. Current activities include: collect best practices and key achievements of each center, define big challenges, and align with other European initiatives. The first meeting of this consortium was in October in Graz, Austria.

Prof. Idris Eckley, Prof. Samuel Kaski, Prof. Geoff Nicholls, Prof. Marina Vannucci Senior Lecturer Veronica Vinciotti,

Scientific Advisory Committee of BigInsight

Scientific Advisory Committee of BigInsight has five international members and met in Oslo in October 2016 for the first time.

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

Prof. Samuel Kaski, University of Helsinki, Finland
- Professor of Computer Science, Aalto University
- Director, Finnish Centre of Excellence in Computational Inference Research COIN, Aalto University and University of Helsinki
- Academy Professor (research professor), 2016-2020
- Statistical machine learning and probabilistic modeling

Prof. Geoff Nicholls, University of Oxford, UK
- Professor in Statistics and Head of Department of Statistics
- PhD in particle physics in the Department of Applied Mathematics and Theoretical Physics in Cambridge, University of Auckland in New Zealand
- Bayesian inference, Computational Statistics, Statistical Genetics, Geoscience, Linguistics and Archaeology

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

Senior Lecturer Veronica Vinciotti, Brunel U of London, UK
- Senior Lecturer in Statistics, Department of Mathematics, Brunel University
- Ph.D in Statistics, Imperial College, London
- Research in statistical classification methods in credit scoring and in statistical genomics
- Co-director of the European Cooperation for Statistics of Network Data Science

ACTIVITIES AND EVENTS

2016 BigInsight Workshop

The yearly BigInsight Workshop was held on Monday 24th October at Telenor Expo, Fornebu. More than hundred researchers and innovators from all BigInsight partners participated to the workshop, during which a wide selection of projects were presented. The Scientific Advisory Committee of BigInsight was present.

The day ended with a happening, namely the last four PhD students who started at BigInsight had a nice cappuccino while chatting about their stories, their expectations, their worries and wishes. Great success for Solveig, Sylvia, Andrea and Emanuele.

2016 Norsk Regnesentral Prize for the best master thesis

Congratulations to Solveig Engebretsen who won the prize for the best master thesis in mathematics and computer science at UiO in 2016. She got the prize for her master thesis titled "Monotone Regression - in high (and lower) dimensions". The prize was split between Solveig and Aslak Wigdahl Bergersen.

Courses

In 2016, BigInsight has organized, supported and sponsored the following courses given at the University of Oslo:

Statistical analysis of high frequency data STK4205

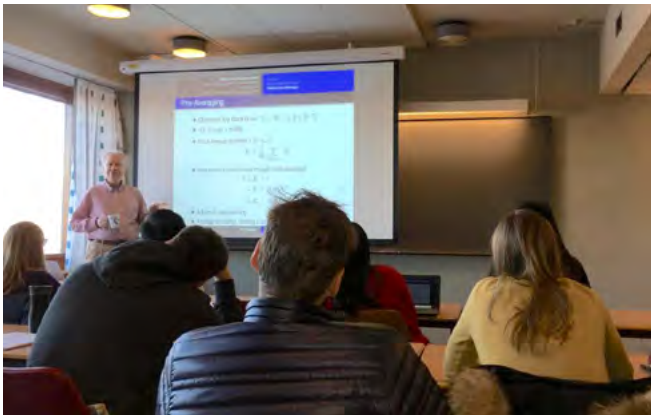
20 students attended the lectures of the intensive course held by Big Insight's visiting professor Per Mykland. The lectures were open to everybody, and we had a nice mix of master and phd students, researchers and partners.

Prediction IMB9275

The first course given in Norway, and among the first world-wide, fully dedicated to statistical predicitons, with molecular biology as main application in mind, but focusing on methodology. The course was part of the Norwegian Graduate School NORBIS in bioinformatics and biostatistics. Lecturers: Arnoldo Frigessi (UiO), Jukka Corander (UiO), Manuela Zucknick (UiO), Ole Winther (Technical Univ. Denmark) and Aki Vehtari (Aalto). PhD students from the whole of Norway participated. The course included a mini-challenge that was part of the exam.

Selected topics in data science (STKINF4000)

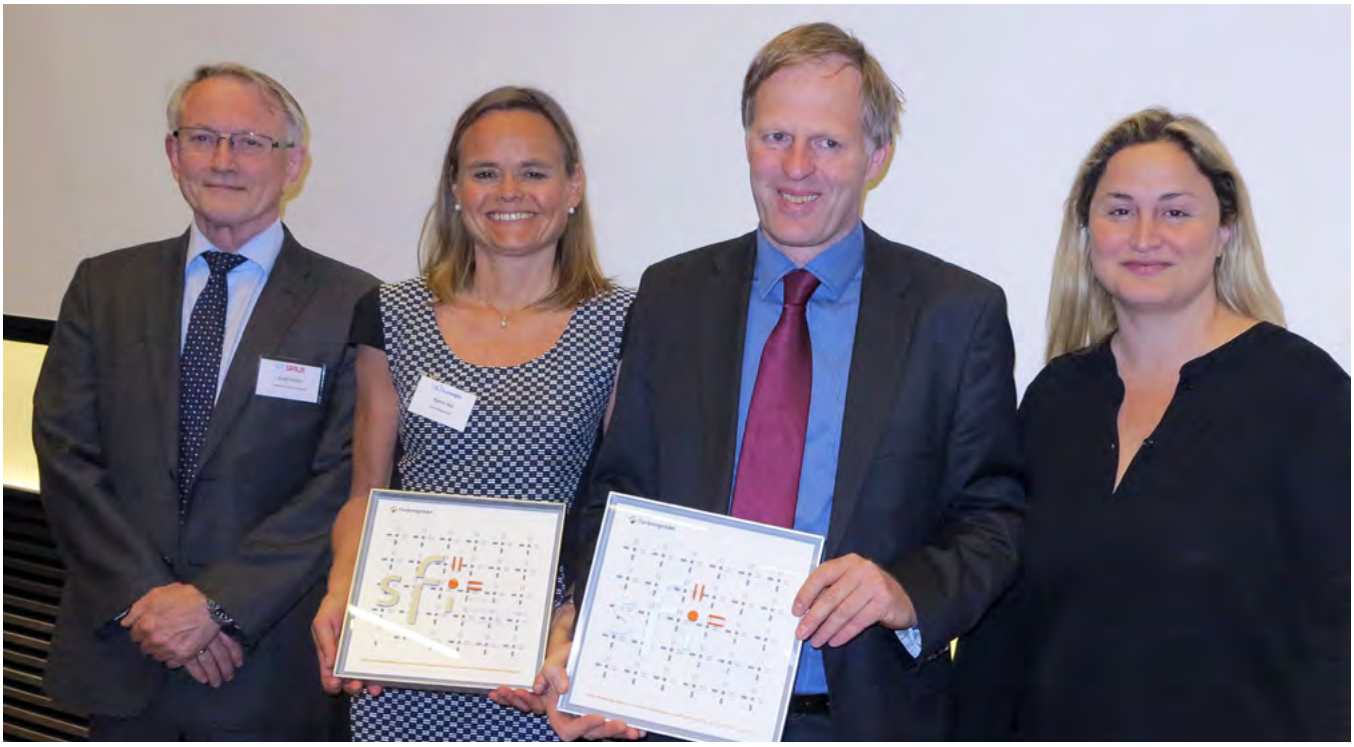
The course bridges gaps between classical statistics and computerscience, in the light of the needs of the industry at large. Applied machine learning, including instruments like Apache SPARK, MongoDB and Scikit-learn. Lectures were given by Dirk Hesse (Intelligent Communication).



Professor Per Mykland

Official Inauguration of BigInsight and SIRIUS

The formal inauguration ceremony of the two centers for research based innovation (sfi) focusing on big data, BigInsight and Sirius was held on Thursday 19th May 2016 by the state secretary from the Minister of Trade and Industry of Norway, Dilek Ayhan and the director of The Research Council of Norway, Arvid Hallén. Both centers received the official sfi plaquette. Congratulations to our cousin SIRIUS!



From left: Arvid Hallén director of The Research Council of Norway, Kjersti Aas codirector of BigInsight, Arild Waaler director of SIRIUS and Dilek Ayhan, state secretary from The Minister of Trade and Industry of Norway. Photo: Bjarne Røsjø, UiO


Moderator: Hans-Olov Adami
Dinner (please note: registration required to attend!)
Henrik Grönberg

COMMUNICATION AND DISSEMINATION ACTIVITIES

Website
biginsight.no is the webpage of the center

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

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

BigInsight in the media
Forskningsnytt, 11.02.2016 «Forskning på statistiske modeller kan gi store gevinster» Arnoldo Frigessi
Dagens Næringsliv, 31.12.2016 «Forbrytelsens matematikk» Anders Løland
Apollon, 16.8.2016, «Ny matematikk skal forutsi farlige sykehusepidemier», Yngve Vogt
forskning.no, 20.1.2016, «Bruk av digitale helseregistre i forskning», Arnoldo Frigessi
ntnu.no, 18.6.2016, «Forskning og Innovasjon – hvordan får vi til begge deler?», Marie H. Aune
sciencenordic.com, 6.9.2016, «New maths to predict dangerous hospital epidemics", Yngve Vogt

The director of Forskningsrådet Arvid Halleen, highlights BigInsight in an innovation conference with Prime Minister Erna Solberg, Thursday, June 2, 2016

BigInsight outreach presentations:

EVENT / **ORGANISER**
Business Analytics Forum / **Sparebank1**
Forskningskafé om Big data / **Forskningsrådet**
Software 2016, Oslo / **Dataforeningen**
Samling for NAV Kunnskap / **NAV**
Skatteetatens analysesamling / **Skatteetaten**
Oslo Life tech workshop / **Oslo Life Tech**
Annual Workshop / **Skatteetaten**
Vårmøte / **Norsk Nyremedisinsk Forening**
Big Health data / **Oslo Life Tech**
Winter seminar My Big Data /
Norsk forening for jus og edb
ESRA seminaret 2016 /
Norsk forening for risiko- og pålitelighetsanalyse (ESRA)
Bedriftsbesøk fra Rauland / **Matnat UiO**
Del og bruk forskningsdata! / **Forskningrådet**
Det nasjonale dekanmøte i medisin / **Dekanmøte i medisin**
Interessentmøte for Innovasjon / **Digital Life Norway**
Lawrence Livermore National Laboratory meeting /
Kreftregisteret
Workshop on the importance
of mathematics for value creation /
Norwegian Mathematical Council
Seminar om cyberphysical systems / **DNV-GL**

RECRUITMENT

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

In 2016 started

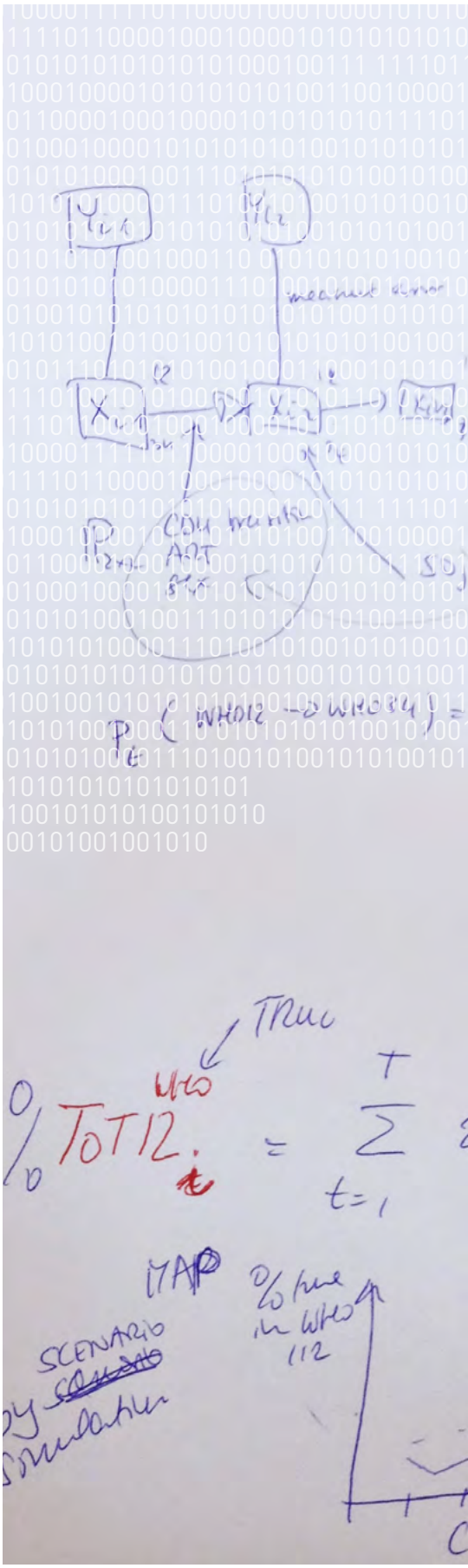
NAME	POSITION	FUNDING	RESEARCH AREA	AFFILIATION
Andrea Chi Zhang	PhD student	BigInsight	Health	UiO, OCBE
Andrea Cremaschi	Postdoc	UiO	Health	Norwegian Centre, Molecular Medicine
Andreas Nakkerud	PhD student	UiO, MI	Power	UiO, Department Mathematics
Andrew Henry Reiner	Researcher	OUS	Health, Marketing	UiO, OCBE
Celia Yinzhi Wang	PhD student	UiO, MI	Fraud	UiO, Department Mathematics
Christian Page	Postdoc	OUS, HSØ	Health	UiO, OCBE
Daniel Piacek	Master student	UiO	Fraud	UiO Department Mathematics
Emanuele Gramuglia	PhD student	BigInsight	Sensor	UiO Department Mathematics
George Zhi Zhao	PhD student	UiO, IMB	Health	UiO, OCBE
Jonas Schenkel	Master student	UiO	Marketing	UiO, Department Mathematics
Jukka Corander	Professor	UiO	Health	UiO, OCBE
Kristoffer Hellton	Researcher	NR	Health	NR
Martin Tveten	Master student	UiO	Sensor	UiO, Department Mathematics
Pierre Lison	Researcher	NR	Marketing	NR
Riccardo De Bin	Associate Professor	UiO	Health, Sensor	UiO, Department Mathematics
Solveig Engebretsen	PhD student	BigInsight & NIPH	Health	UiO, OCBE and NIPH
Sylvia Qinghua Liu	PhD student	UiO, MI innovation	Health	UiO, Department Mathematics
Vinnie Ko	PhD student	UiO	Fraud	UiO, Department Mathematics

PERSONNEL

Personnel affiliated with BigInsight

NAME	INSTITUTION	MAIN RESEARCH AREA
Rune Braastad	ABB	Sensor
Stian Braastad	ABB	Sensor
Kenneth Nakken	ABB	Sensor
Jaroslaw Novak	ABB	Sensor
Gunnar Prytz	ABB	Sensor
Bettina Kulle Andreassen	CRN	Health
Bjørn Møller	CRN	Health
Jan Nygård	CRN	Health
Hege Bolsø	DNB	Marketing
Thorstein Danielsen	DNB	Fraud
Thor Aage Dragsten	DNB	Marketing
Heidi Fredriksen	DNB	Marketing
Katharina Henriksen	DNB	Marketing
Roy Oma	DNB	Fraud
Steffen A. S. Sjursen	DNB	Marketing
Fredrik Strand	DNB	Fraud
Mette S. J. Snilsberg	DNB	Marketing
Geir Ånonsen	DNB	Fraud
Øystein Alnes	DNV-GL	Sensor
Theo Bosma	DNV-GL	Power
Ervin Bossanyi	DNV-GL	Power
Frédéric Courivaud	DNV-GL	Health
Marcel Eijgelaar	DNV-GL	Power
Odin Gramstad	DNV-GL	Sensor
Muhammad Jafar	DNV-GL	Power
Lars Landberg	DNV-GL	Power
Stephen Leyshon	DNV-GL	Health
Gabriele Manno	DNV-GL	Sensor
Bahman Raeissi	DNV-GL	Sensor
Bobby Ray-Sannerud	DNV-GL	Health
Elizabeth Traiger	DNV-GL	Power
Erik Vanem	DNV-GL & UiO	Power, Sensor
Bjørn-Johan Vartdal	DNV-GL	Sensor
Tone Sejnæs Eilertsen	Gjensidige	Marketing
Mikkel Hinnerichsen	Gjensidige	Fraud, Marketing
Randi Jule	Gjensidige	Fraud
Anders Nyberg	Gjensidige	Marketing
Marte Olstad	Gjensidige	Marketing
Geir Inge Sandnes	Gjensidige	Marketing
Thea Margrethe Skouen	Gjensidige	Marketing
Geir Thomassen	Gjensidige	Fraud
Stefan Erath	Hydro	Power
Plamen Mavrodiev	Hydro	Power
Ivar Areklett	NAV	Fraud
Eigil Johnsen	NAV	Fraud

NAME	INSTITUTION	MAIN RESEARCH AREA
Bjørn Atle Sundsback	NAV	Fraud
Morten Tholander	NAV	Fraud
Jonas Andersson	NHH	Fraud
Birgitte De Blasio	NHI	Marketing, Health
Kjersti Aas	NR	Marketing, co-director, PI
Magne Aldrin	NR	Sensor, PI
Line Eikvil	NR	Fraud
Clara-Cecilie Günther	NR	Health, Marketing, PI
Martin Jullum	NR	Fraud
Ola Haug	NR	Marketing, Sensor
Marion Haugen	NR	Marketing
Kristoffer Hellton	NR	Health
Lars Holden	NR	Health
Marit Holden	NR	Health
Ragnar Bang Huseby	NR	Power, Fraud, PI
Alex Lenkoski	NR	Power, PI
Pierre Lison	NR	Marketing
Anders Løland	NR	Fraud, Power, PI
Linda R. Neef	NR	Fraud
Hanne Rognebakke	NR	Marketing, Sensor
Nikolai Sellereite	NR	Marketing
Gunnhildur Steinbakk	NR	Fraud, Power
André Teigland	NR	Co-director
Ingunn Fride Tvete	NR	Health
John Enok Vollestad	NR	IT
Tor Arne Øigård	NR	Power, Fraud
Mette Langaas	NTNU	Health
Ragnhild Sørum Falk	OUS	Health
Harald Fekjær	OUS	Health
Jon Michael Gran	OUS	Marketing, Health
Eivind Hovig	OUS	Health
Marissa LeBlanc	OUS	Health
Sygve Nakken	OUS	Health
Andrew Henry Reiner	OUS	Health, Marketing
Anders Berset	Skatteetaten	Marketing, Fraud
Wenche Celiussen	Skatteetaten	Marketing, Fraud
Øystein Olsen	Skatteetaten	Marketing
Audun Solli	Skatteetaten	Marketing, Fraud
Fatima Yusuf	Skatteetaten	Marketing
Geoffrey Canright	Telenor	Marketing, Health
Humberto Castejon	Telenor	Marketing
Kenth Engo-Monsen	Telenor	Marketing, Health
Anita Fjuk	Telenor	Marketing
Gro Nilsen	Telenor	Marketing
Hai Thanh Nguyen	Telenor	Marketing
Massimilano Ruocco	Telenor	Marketing
Astrid Undheim	Telenor	Marketing
Jan Bulla	UiB	Marketing
Bård Støve	UiB	Fraud
Dag Tjøstheim	UiB	Marketing, Sensor
Elja Arjas	UiO	Marketing, Health



NAME	INSTITUTION	MAIN RESEARCH AREA
Ørnulf Borgan	UiO	Marketing
Jukka Corander	UiO	Health
Arnoldo Frigessi	UiO	Marketing, Health, Sensor, PI
Ingrid K. Glad	UiO	Sensor, co-director, PI
Ingrid Hobæk Haff	UiO	Fraud, co-director, PI
Arne Bang Huseby	UiO	Sensor, Power
Carlo Mannino	UiO	Power, Sensor, PI
Kjetil Røysland	UiO	Health
Geir Kjetil Sandve	UiO	Health
Ida Scheel	UiO	Marketing
Geir Storvik	UiO	Sensor
Magne Thoresen	UiO	Health
Marit Veierød	UiO	Health
Manuela Zucknick	UiO	Health
Sylvia Richardson	BSU-MRC	Health
Idris Eckly	Lancaster	Sensor

NAME	FUNDING	NATIONALITY	PERIOD	GENDER	TOPIC
Postdoctoral researchers in BigInsight with financial support from other sources					
Valeria Vitelli	UiO/Kreftforening/NFR	Italy	2015-2020	F	Marketing, Health
Kristoffer Hellton (20%)	UiO/NFR Focustat	Norway	2015-2017	M	Health
Gudmund Hermansen	UiO/NFR Focustat	Norway	2016-2018	M	Sensor
Christian Page	OUS/HSØ	Norway	2016-2018	M	Health
Andrea Cremaschi	UiO, NCMM	Italy	2016-2018	M	Health
PhD students with financial support from BigInsight					
Håvard Kvamme	Norway	2015-2019	M	Marketing	
Andrea Chi Zhang	China	2016-2018	F	Health	
Solveig Engebretsen	Norway	2016-2018	F	Health	
Emanuele Gramuglia	Italy	2016-2018	F	Sensor	
PhD students in BigInsight with financial support from other sources					
Derbachew Asfaw	UiO/Norhed	Ethiopia	2015-2016	M	Marketing
Marta Crispino	Univ Bocconi Milano	Italy	2015-2016	F	Marketing
Xiaoran Lai	UiO	UK	2015-2018	M	Health
Andreas Brandsæter	DNV-GL, NFR NæringsPHD	Norway	2015-2018	M	Sensor
Jamie-Leigh Chapman	Univ Lancaster	UK	2015-2017	F	Sensor
Andreas Nakkerud	UiO/Dep. Mathematics	NO	2016-2018	M	Power
Celia Yinzhi Wang	UiO, Dep. Mathematcis	China	2016-2018	F	Fraud
George Zhi Zhao	UiO, IMB	China	2016-2018	M	Health
Sylvia Qinghua Liu	UiO MI innovation	China	2016-2018	F	Health
Vinnie Ko	UiO	Dutch	2017-2018	M	Fraud
Master degrees					
Martin Tveten				M	Sensor
Jonas Schenkel				M	Marketing
Daniel Piacek				M	Fraud

FINANCIAL OVERVIEW

FUNDING	1000 NOK
The Research Council	8 259
Norwegian Computing Center (NR)	1 558
Research Partners*, in kind	8 706
Research Partners*, in cash	468
Enterprise partners**, in kind	5 884
Enterprise partners**, in cash	4 097
Public partners***, in kind	4 268
Public partners***, in cash	1 343
Sum	**34 583**

COSTS	
NR, research	11 314
NR, direct costs	778
Research Partners*, research	11 939
Enterprise partners**, research	5 884
Public partners***, research	4 668
Sum	**34 583**

* Research partners: UiO, UiB

** Enterprise partners: Telenor, DnB, Gjensidige, Norsk Hydro, DNV-GL, ABB

*** Public partners: Norwegian Tax Administration (Oslo), University Hospital HF, NAV, Public Health Insitute (NIPH)

PUBLICATIONS IN 2016

(First papers of BigInsight start to appear in print on scientific peer reviewed journals and conference proceedings.)

Peer reviewed scientific papers

Aas, Kjersti. "Pair-Copula Constructions for Financial Applications: A Review." Econometrics 4.4 (2016): 43.

Bolin, David; Frigessi, Arnoldo; Guttorp, Peter; Haug, Ola; Orskaug, Elisabeth; Scheel, Ida; Wallin, Jonas.
Calibrating regionally downscaled precipitation over Norway through quantile-based approaches. Advances in Statistical Climatology, Meteorology and Oceanography 2016 ;Volum 2. s.39-47

Brandsæter, Andreas; Manno, Gabriele; Vanem, Erik; Glad, Ingrid Kristine.
An Application of Sensor-Based Anomaly Detection in the Maritime Industry. I: 2016 IEEE International Conference on Prognostics and Health Management (ICPHM 2016). Curran Associates, Inc. 2016 ISBN 9781509003839.

Frigessi, Arnoldo; Vitelli, Valeria; Crispino, Marta; Arjas, Elja.
Recommendations from non-transitive pairwise comparisons. RecSys 2016; 2016-09-12

Glad, Ingrid Kristine; Hjort, Nils Lid.
Model uncertainty first, not afterwards. Statistical Science 2016 ;Volum 31.(4) s.490-494

Hobæk Haff, Ingrid; Aas, Kjersti; Frigessi, Arnoldo; Lacal Graziani, Virginia.
Structure learning in Bayesian Networks using regular vines. Computational Statistics & Data Analysis 2016 ;Volum 101. s.186-208

Solbergersen, LInn Cecilie; Ahmed, Ismail; Frigessi, Arnoldo; Glad, Ingrid Kristine; Richardson, Sylvia Therese Lamblin.
Preselection in Lasso-Type Analysis for Ultra-High Dimensional Genomic Exploration. Abel Symposia 2016 ;Volum 11. s.37-66

Tharmaratnam, K.; Sperrin, M.; Jaki, T.; Reppe, Sjur; Frigessi, Arnoldo.
Tilting the lasso by knowledge-based post-processing. BMC Bioinformatics 2016 ;Volum 17.(344) s.1-9

Vanem, Erik; Brandsæter, Andreas; Gramstad, Odin.
Regression models for the effect of environmental conditions on the efficiency of ship machinery systems. I: Risk, Reliability and Safety: Innovating Theory and Practice : Proceedings of ESREL 2016 (Glasgow, Scotland, 25-29 September 2016). CRC Press 2017 ISBN 9781138029972.

Reports

Aas, Kjersti; Rognebakke, Hanne Therese Wist.
Analysis of Vipps data. Norsk Regnesentral 2016 24 s. NR-notat(SAMBA/20/16) NR

Holden, Lars.
The two subset recurrent property of Markov chains. Norsk Regnesentral 2016 25 s. NR-notat(ADMIN/01/2016) NR

BigInsight

biginsight.no

Postal address
BigInsight
Pbox 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@nr.no
Anders Løland anders@nr.no

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

Design: Designavdelingen, Kunst i Skolen


BigInsight