Departement Statistiek en Aktuariële Wetenskap
Department of Statistics and Actuarial Science

Nuusbrief | Newsletter
December 2019


PAGE 8


PAGE 15


PAGE 17

IN THIS ISSUE

BIG COMPUTER UPGRADE FOR BIG DATA

In the age of big data, a lot of emphasis is placed on buying ever more powerful computers. These computers are used to record and analyse data and to run simulation studies to predict future outcomes. To this end the Department of Statistics and Actuarial Science has invested in two server-level workstations to ensure that staff, researchers and students have access to the necessary computer resources to keep abreast of big data software, handling and simulation. **Dr Chris Muller** tells us more…

More powerful has most often come to mean computers with faster processors (although today more powerful tends to mean a higher number of processors). Processors (CPUs – Central Processor Units) are the heart of a computer, enabling it to execute a program and to complete a specified task. The speed of a processor (known as the clock-speed and represented by a number in GHz) determines how quickly a processor will be able to complete a given task. These speeds have increased from the 0.45 to 0.6 GHz range of the early 2000s to the 4 to 5 GHz range of today.

Not only have CPUs become faster, they have also evolved from being able to execute a single task to being able to execute multiple tasks simultaneously (parallel computing). This is possibly due to a single physical processor now having multiple cores. Each core is in effect a processor that can execute programs and tasks, independently from the other cores. A single processor computer with eight cores can run eight programs simultaneously while a dual core processor (two physical processors with eight cores each installed on one motherboard within one physical computer body), can run 16 programs simultaneously. Previously, this would have required 16 physical computers. If used correctly, i.e. using software that can utilise all the cores at once and writing code that can split a program into smaller parts, this means that a simulation that used to take sixteen hours can be completed in around one hour on this computer.

A newer type of processing unit, the graphics processing unit (GPU), has started playing an ever-increasing role in data warehousing, machine learning, artificial intelligence and big data manipulation and analysis. These "graphics cards" (the colloquial term for a GPU) are generally used for very mathematically intensive tasks associated with three-dimensional (3D) rendering and modelling. Here complicated tasks require thousands of basic tasks to be performed. Each time a 3-D object is moved thousands of calculations need to be performed to ensure that the object being moved keeps its shape. Each calculation is basic in nature, but all must be completed at the same time to make sure that the object retains its correct shape as it is rotated on the screen. Using the CPU to perform these thousands of calculations in series (completing one before starting the next one) would bring a computer to a standstill every time a three-dimensional object is viewed. Unlike CPUs with eight or 16 cores, a GPU can have up to 5,000 cores. These cores allow a GPU to complete thousands of calculations simultaneously and literally get the 3-D ball smoothly rolling down the hill without bringing a computer to a standstill.

Computer programmers have caught on to the fact that these thousands of slower cores (0.9 GHz to 1.6 GHz) on a GPU, usually sitting idle, can be used not only for animation and 3-D modelling, but also for number crunching. By specifically writing a program to make use of the massive parallel nature of a GPU, tasks can be completed significantly faster than using fewer but faster CPU cores. Due to this, more and more companies are, rather than upgrading their CPUs and gaining 16 x 4.5 GHz cores, installing specialised GPUs and gaining 10 000 x 1.4 GHz cores.

Memory (RAM, random access memory) is an often-overlooked quantity when buying a new computer. Today's computers have as standard configuration 8GB of RAM and there are no real gains for most users in upping this to 16GB or more. This all changes when analysing big data. It is possible to analyse big data on modern super-fast 5 GHz 32-core computers, as well as archaic 1.5 GHz single-core computers. If the operating system and simulation software will run on a computer, irrespective of the clock-speed or the number of cores, a simulation study will run. On the old computer it will be an extremely frustrating experience and it may take weeks or months, but a researcher will get results. Unlike having an old computer that will just take time to get results, not having enough memory for a big data simulation or analysis, be that on an old or a brand-new computer, will stop a big data simulation or analysis dead in its tracks.

Due to not being able to run a big data simulation or analysis without enough memory, memory has become even more important for big data than the clock-speed of a processor or the number of cores. Consumer-level computers have only very recently been upgraded to accommodate 64GB of memory and are therefore not appropriate for handling big data problems. Big data necessitates the use of server-level processors that can handle up to 768GB of memory.

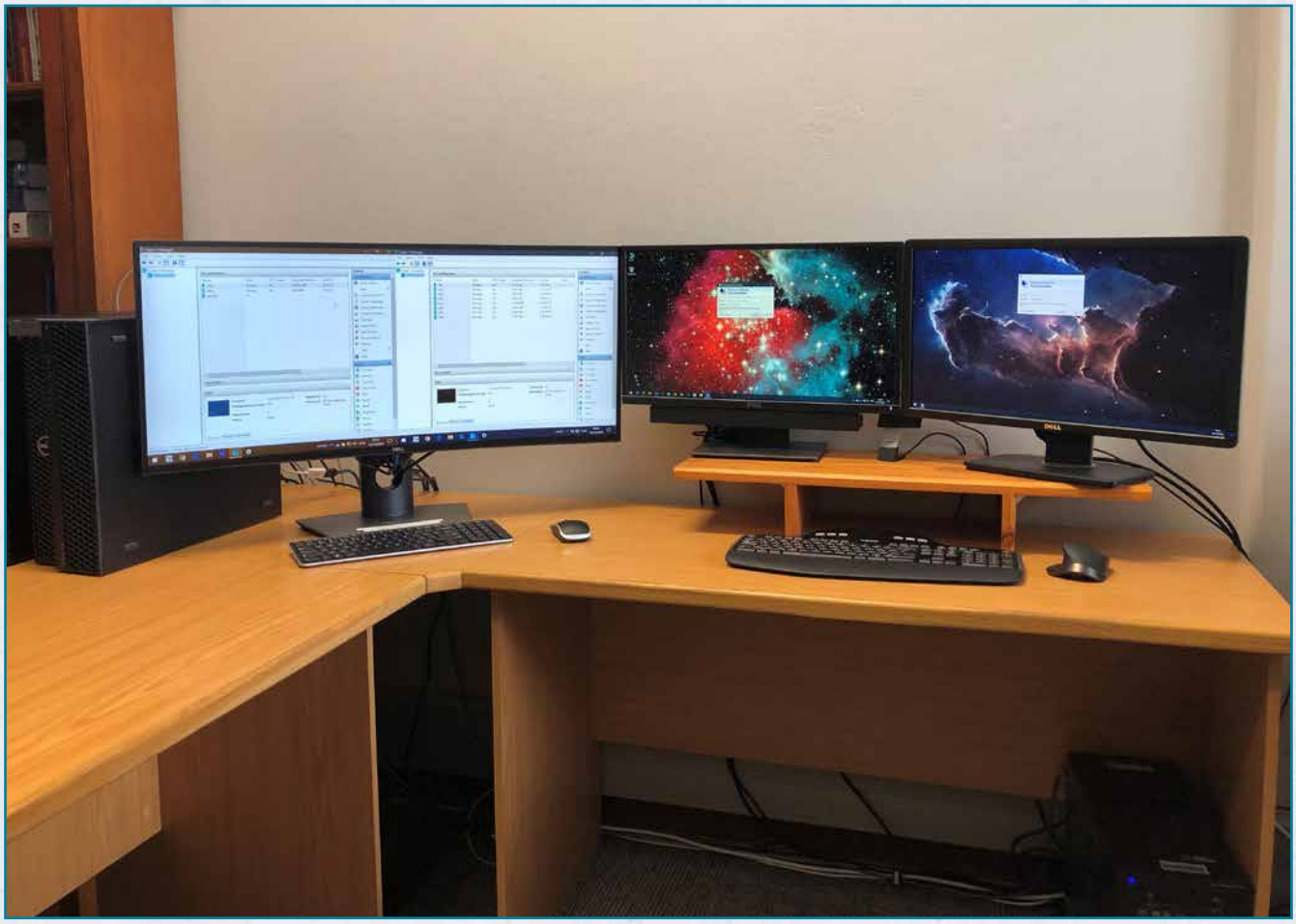
BIG COMPUTER UPGRADE FOR BIG DATA

To this end the Department has invested in two workstations with server-level components. The Department's two workstations boast 72 cores running at 3.7 GHz, 24TB of storage space (4TB on fast solid-state drives), 768GB RAM and four high-end nVidia workstation GPUs that can perform up to 8.9 TFLOPs (8.9 trillion floating-point operations per second).

Virtualisation software is utilised to create virtual machines (VMs) for researchers and students. This allows for the customisation of the operating system (Linux or Windows), number of processors, the amount of memory, and the storage space. For example, a student with an old and slow laptop may require a VM with only two processors, 8GB RAM and 500GB of storage to run a basic simulation, while a researcher with a big data set may require 30 processors, 250GB RAM and 4TB of storage. Each VM is a fully-fledged computer and the different VMs and users are not aware of the fact that they are sharing the physical computer and its resources. Using VMs the scenario where a computer stands in one individual's office and when that person is not using it, it is sitting idle, is avoided. It also has the advantage that if a simulation goes wrong, it is at most a 30-minute exercise to delete that specific VM and create a new VM using preconfigured templates.

Scheduling software is used to show how the different VMS are being utilised, allowing users to quickly see which VMs are being used, when the next VM will be free, and the specifications for the different VMs. Since being put into production there were times that up to 12 researchers or students were using the workstations at a time.

The investment in these two workstations ensures that staff, researchers and students have access to the necessary computer resources to keep abreast of big data software, handling and simulation. This will also allow students that will follow the new Bachelor of Data Science (BDatSci) programme from 2021 onwards to have practical exposure to handling big data on server-level computers.



Campanology research uses Statistics to tell the story of bells

Prof Tertius de Wet of the Department of Statistics and Actuarial Science – together with Prof Jef Teugels from the Katholieke Universiteit Leuven and a departmental colleague, Dr Pieta van Deventer – have been doing research on bells in the Western Cape since 2008. The study of bells and bellringing is referred to as Campanology. **Prof De Wet** tells us more…

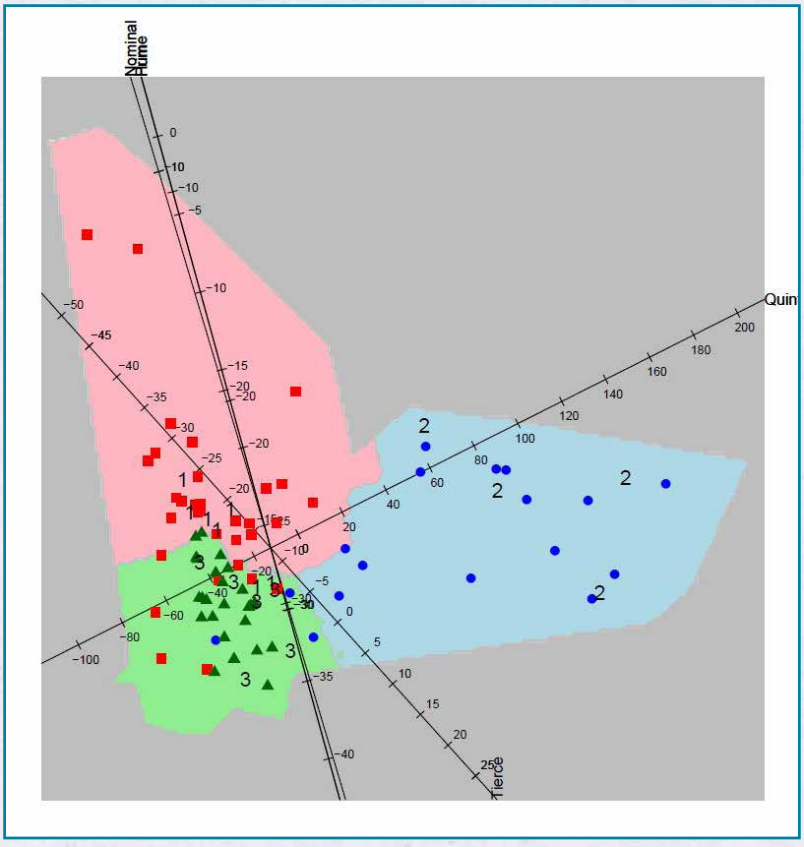
Research projects were initiated to locate bells in the Western Cape, especially those of historical value, and to gather as much information as possible on them, through visits and/or documents and other material.

For us this has been and still is an exciting field of research. Apart from the historical aspects, we also analyse data obtained from measuring the physical and acoustical properties of the bells. We refer to this part of our Campanology research as *Campanometry*, the statistical study of bells.

Since 2008 we have succeeded in collecting information on more than 450 bells. The modus operandi is to visit a bell(s) in its current environment, take pictures of it, measure its physical dimensions (height, diameter and thickness at the sound bow), record its sound and to note all possible decorations and inscriptions on the bell. The sound recordings are analysed to find the bell's five main overtones and these are statistically compared to its intended notes.

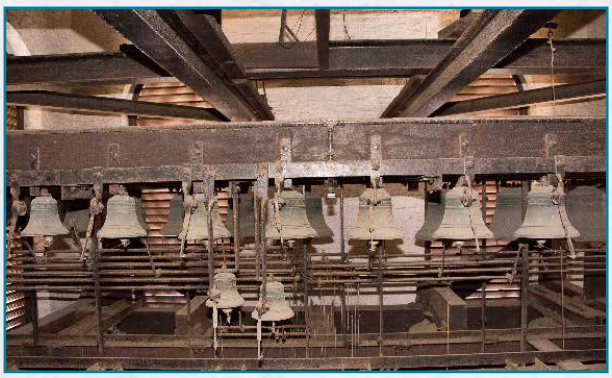
These analyses include exploratory analyses, model fitting of deviations from intended values of overtones, analysing and comparing the sound quality of bells and carillons, and founder recognition using statistical classification. The latter aspect is especially relevant to bells in the Western Cape where a large number are either undated and/or for which the founder is unknown. This calls for some challenging classification applications, using the data available as well as bell shape and other characteristics unique to a particular bell founder.

A biplot is one method useful for classification, especially to visually present the data, variables and classes. As an example, in the figure on the right a biplot gives the classification of a set of bells into one of three categories of founders: Taylor (red area), Eijsbouts (blue area), and Gillett & Johnston (green area). Training data are represented by shapes: Taylor (red squares), Eijsbouts (blue circles), and Gillett & Johnston (green triangles). Testing data are represented by numbers: Taylor (1s), Eijsbouts (2s), and Gillett & Johnston (3s). The axes denote the variables, four overtones in this case.



One of the first projects carried out was on the 39-bell carillon in the Cape Town City Hall. This instrument is one of only two manually playable carillons on the African Continent – the other being in the campanile in Port Elizabeth. The carillon was one of the first instruments worldwide to be conceived as a memorial for the victims of the First World War (1914-1918).

STATISTICS IN CAMPANOLOGY RESEARCH



The picture above on the left shows the top layer of smaller bells of the carillon and the one on the right the bell in honour of those who lost their lives in the battle in Delville Wood, one of the worst battles for South African soldiers in World War I.

An important aspect of a carillon is the quality of its sound. One such measure is a weighted average of deviations from their intended values of the four lowest overtones, measured against threshold values derived from campanology theory. For the Cape Town carillon, its value for this measure was calculated and compared to the value of the 63-bell carillon in the university library of the Katholieke Universiteit Leuven. As a measure of uncertainty in these estimated quality values, confidence intervals are useful for assessing differences between the obtained values. Since distributional assumptions are unrealistic to make, a nonparametric bootstrap resampling method was used to obtain 95% confidence intervals for the mean quality measure of each carillon. These intervals were found not to overlap with the Cape Town carillon having a lower mean and a tighter confidence interval, showing it to be of higher quality than the Leuven one – see the table below.

	Cape Town	Leuven
Mean	44.51	59.40
Std. error	2.00	2.81
Lower Bound	40.59	54.02
Upper Bound	48.45	65.29

The latter is especially significant given that the Cape Town carillon has been hanging in the City Hall bell tower, close to the ocean, for 90+ years – it is well known that close proximity to sea air over time could have a negative effect on the sound quality of bells.

Many of these bells we research are located in church towers, often leading to challenging climbs to reach the bells for data collection, illustrated in the pictures below. These show the church tower of the Eendekuil Dutch Reformed Church on the left and the climb to reach the bells on the right. A high-adrenalin climb!

STATISTICS IN CAMPANOLOGY RESEARCH

Researching historical bells often entails going through old documents in archives, in itself a very interesting activity, but sometimes you unexpectedly come upon a historic bell. For example, in March 2018 we were collecting data on a number of bells known to us in the Constantia area of Cape Town, when we stopped to have lunch at the Alphen Hotel. Great was our surprise when we saw a lovely bell hanging in a horseshoe-type bell tower, a clearly historical bell and one totally unknown to us. On the bell is found the following inscription in Latin:

ME FECERUNT AMSTELODAMI A° 1705 + CLAES NOORDEN ET IAN DE GRAVE

This indicates that the bell was founded in Amsterdam in 1705 by Claes Noorden and Jan de Grave. This is indeed a bell of historical value – see the pictures below.

Although this is quite an old bell for South Africa, it is not the oldest one we have found to date.

That honour belongs to one of the bells of the Stellenbosch Rhenish Church, a bell dated 1664, a rare example of the founder Johan Dop who worked in Utrecht, Netherlands, and of whom only a few of his bells still exist.

The collection of bell information is ongoing. This information obtained is stored on the Stellenbosch University Library's digital platform at https://digital.lib.sun.ac.za/10019.2/3975

On this website the information is available to other researchers and to the general public. Over the years a number of questions and comments were received from people who have consulted the website. The website is also expanded from time to time as more information becomes available. We use the information as a basis for writing academic papers, both of a historical nature as well as of a technical nature, and also more popular papers. Furthermore, the existence of such an inventory will hopefully in some way also assist in preventing (further) disappearance of this part of our cultural heritage.

For us, this is an exciting field to be involved in: to preserve and record some of the micro history of the Western Cape and to develop another area of application of Statistics. Due to logistical reasons, our current research is confined to the Western Cape. However, we know of a number of historic bells in other provinces which would be fruitful areas for further research.

STAFF MATTERS

Research showcased internationally

The 16th biennial IFCS conference took place in Thessaloniki, Greece, from 26-29 August 2019. Delegates from 29 countries attended the conference; among the 224 delegates were eight participants from South Africa. Six of the eight South African delegates represented Stellenbosch University.

Prof Niël le Roux, emeritus professor in the Department of Statistics and Actuarial Science, organised two invited-paper sessions on behalf of the multivariate data analysis group (MDAG), a special interest group of the South African Statistical Association (SASA) and an affiliated member of the International Federation of Classification Societies (IFCS).

These sessions were titled "Classification, visualisation and dimension reduction" and dedicated to the late Prof John Gower. The speakers (in bold) in the MDAG session were:

- **Prof Rénette Blignaut** (UWC), Prof Isabella Venter (UWC), Dr Humphrey Brydon (UWC)
  Topic: Using separate sampling to understand mobile phone security compliance.
- **Dr Renato Cordeiro de Amorim** (University of Essex)
  Topic: Unsupervised feature selection and big data.
- **Mrs Raeesa Ganey** (Wits), Prof Sugnet Lubbe (SU)
  Topic: Visualising multivariate data in a principal surface biplot.
- **Prof Martin Kidd** (SU)
  Topic: A multivariate ROC based classifier.
- **Prof Niël le Roux** (SU), Prof John Gower
  Topic: Properties of individual differences scaling and its interpretation.
- **Prof Sugnet Lubbe** (SU)
  Topic: Functional linear discriminant analysis for several functions and more than two groups.
- **Dr Johané Nienkemper-Swanepoel** (SU), Prof Niël le Roux (SU), Prof Sugnet Lubbe (SU)
  Topic: A simulation study for the identification of missing data mechanisms using visualisations.
- **Dr Trudie Sandrock** (SU)
  Topic Local and global relevance of features in multi-label classification.
- **Prof Danie Uys** (SU)
  Topic: Tree-based ensemble methods for classification.

The next IFCS conference will be held at the University of Porto, Portugal, from 25-28 August 2021.

The speakers in the MDAG session were from left to right Johané Nienkemper-Swanepoel, Martin Kidd, Rénette Blignaut, Trudie Sandrock, Danie Uys, Sugnet Lubbe, Renato Cordeiro de Amorim, Isabella Venter and Niël le Roux.

STAFF MATTERS

Department hosts workshop on Bayesian Biostatistics

The Department hosted a very successful workshop on Bayesian Biostatistics in November, presented by Prof Emmanuel Lesaffre. Prof Lesaffre is an emeritus professor from the Katholieke Universiteit Leuven who regularly visits the Department for research collaboration. His visit to Stellenbosch University (SU) was partly sponsored by the UCDP grant of DST and the NRF. The participants came from a number of Departments of Statistics in South Africa, such as University of the Western Cape, University of KwaZulu-Natal, North West University, University of Zululand, University of Pretoria, University of the Free State and Stellenbosch University. Young academics were fully sponsored by the UCDP to attend this workshop. The funding application to secure their attendance was done by Dr Chris Muller and Prof Paul Mostert of SU's Department of Statistics and Actuarial Science. Pictured here is a group photo of all the delegates with Emmanuel Lesaffre and Chris Muller.

Congratulations to Justin, Willie and Luca

Dr Justin Harvey recently received a C3 rating from the NRF after submitting his application earlier this year. The C category describes a person as an established researcher with a sustained recent record of productivity in the field who is recognised by their peers as having produced a body of quality work, the core of which has coherence and attests to ongoing engagement with the field. This person has also demonstrated the ability to conceptualise problems and apply research methods to investigating them.

The 9th edition of the top lecturer competition sponsored by *Die Burger* took place in September this year. This year, students voted for the top undergraduate and postgraduate lecturers in each of the seven departments of the Faculty of Economic and Management Sciences. Congratulations to Mr Luca Steyn and Prof Willie Conradie for being voted the top undergraduate and postgraduate lecturers in our Department, respectively.

Dr Justin Harvey

Mr Luca Steyn

Prof Willie Conradie

Denver's passion for soccer

Denver Loggenberg, an admin assistant in the Department of Statistics and Actuarial Science, has been passionate about soccer since he was a young boy. In 2018, he joined the Stellenbosch University (SU) staff soccer team and in 2019, he represented Stellenbosch University at the University Sport South Africa (USSA) national staff tournament in Durban.

Where did your interest in soccer begin?

I started playing soccer for Nelson's Football Club in Idas Valley, Stellenbosch, at the age of 10. When I went to high school, I played for Luckhoff High's soccer team. When I was about 18, I started playing for United Football Club, also in Idas Valley. They noticed my soccer talent while I was still playing for Luckhoff. I later joined Idas Valley Football Club, around 2014, where I'm still playing soccer today. In 2017 I briefly played for Maties Football Club, for about two seasons, but I returned to Idas Valley Football Club in 2018. I now represent both Idas Valley Football Club and the SU staff soccer team.

Denver (green and white) representing Idas Valley Football Club

Tell us about your soccer highlights

When I was still at Luckhoff High, we won an important game against Cloetesville High School. For this game I was man of the match, which inspired me to continue further with my soccer career. One year, while playing for Idas Valley Football Club, we won the Castle league and were promoted to the Vodacom league. This was a very special moment for us as a team. On 8 October this year, the US staff soccer team won against the UWC staff soccer team in a match played here in Stellenbosch. I scored three of our four goals in this game, a very special highlight for me. On 25-27 October we participated in the USSA national staff tournament in Durban. Staff teams from 14 different universities competed in this tournament. In the group stage we played against Durban University of Technology (DUT), Nelson Mandela University (NMU) and Sol Plaatje University (SPU). We drew 2-2 against DUT, a match in which I scored one of the goals. We also drew against NMU and lost against SPU. Unfortunately, we did not make it pass the group stage, but I still consider this tournament a special occasion in my soccer career.

What are your other interests?

I like to follow the local soccer as well as international soccer games. Locally I support Stellenbosch Football Club and internationally I support Manchester United. I enjoy working at the Department and at home I enjoy spending time with my family, especially my young daughter.

News from the 2019 SASA conference

Professor Paul Mostert was awarded Fellowship from the South African Statistical Association (SASA). (A member who has made outstanding contributions to the advancement of Statistical Science may be nominated and elected as a Fellow of the Association.)

Professor Daan Nel was awarded Honorary membership from SASA. (An Honorary member is someone who has made an extraordinary contribution to the knowledge of Statistics, or to the advancement of the association, and who has retired from his/her ordinary position.)

STAFF MATTERS

Professor represents SA on international golfing stage

Prof Garrett Slattery (captain of the Rest of the World team), Chris Hilton (captain of the Royal and Ancient Golf Club of St Andrews) and Linda Port (captain of the North American team).

Prof Garrett Slattery, head of Actuarial Science, was selected as playing captain of the Rest of the World team for the ISPS Handa Vision Cup (the International Blind Golf Association's version of the Ryder Cup) earlier this year. The event was held at Portmarnock Links near Dublin, Ireland. This was the 4th staging of the event, the last one in Canada in 2017 having ended in a 12-12 tie. Garrett and his Italian partner won their opening two matches, and the Rest of the World team went into the final day with a three-point lead over the North Americans. A strong performance in the singles matches on the final day (including Garrett's win over the 2019 US Open champion) resulted in victory for the Rest of the World team by 15.5-8.5 points.

In September, Garrett participated in the ISPS Handa Japan Blind Open, held at the picturesque Hakone Golf Club not far from Tokyo. Despite the nearby volcano letting off steam, and some heavy rain on the first day, the event went off well. Garrett finished in 2nd place gross in a field of over 50 players. The following week Garrett successfully defended his SA title at the Glendower Golf Club in Johannesburg.

Staff milestones

Nappy drills lead to new challenges: On 20 September, Remiel (Remi) was born to senior lecturer Martin Coxon and his wife Rickma. Remi made a peaceful journey from womb to world and both mother and son are doing well post-partum. While Remi (surprisingly) lets his parents get some sleep - he insists on putting them through their paces when it comes to nappy changes, especially his dad. On behalf of the Department we wish Martin, Rickma and Remi all the best as a family.

Remi Coxon

Department bids farewell to Martin and Rousseau: On 1 November, we said goodbye to two of our colleagues, Martin Coxon and Rousseau Lötter. Martin joined the Department in 2017 as a senior lecturer in Actuarial Science while Rousseau joined the Department in 2018 as a senior lecturer in Financial Risk Management. Even though their time at the Department was short, we trust that they have enjoyed this time. On behalf of the Department we would like to wish them well with their future plans and professional careers.

Mr Martin Coxon

Dr Rousseau Lötter

Financial Risk Management students win Hackathon data science challenge

Daniel Bugan

Students from the Department of Statistics and Actuarial Science came out tops in the data science challenge section of Stellenbosch University's second Hackathon.

The annual event, which ran from 2-4 August at SU's Engineering main building, was organised by Innovus (the university industry interaction and innovation company of Stellenbosch University (SU)) and sponsored by Entersekt, Capitec and Explore. The Hackathon aims to increase an entrepreneurial and innovation culture across the SU campus by getting talented student programmers, designers, builders, statisticians, scientists and engineers to come together to learn, build and share their creative ideas, while solving industry-relevant problems in an innovative way.

For the first time, students could participate in one of two challenges: a data science challenge or a financial technology (fintech) coding challenge. The fintech challenge was sponsored and set up by Capitec and Entersekt, with help from the Department of Electrical and Electronic Engineering and the Department of Mathematical Sciences. The data science challenge was sponsored and set up by Capitec and Explore with help from the Department of Statistics and Actuarial Science. The theme of both challenges was to design useful and affordable products and services for consumers that are under-served and financially excluded. The data science challenge required teams to use insights obtained from client transactional data, and to suggest the next best action or tip based on what the team had learnt from the data science or machine-learning process. The team consisting of David Rodwell, Christiaan le Roux and Robert Cronjé – all Financial Risk Management (FRM) master's students in the Department of Statistics and Actuarial Science – came out tops. They were the most successful in using their creative problem-solving, data-cleaning, manipulation and visualisation skills, as well as their experience in statistical analysis.

Mr Carel van der Merwe, senior lecturer in FRM at the Department of Statistics and Actuarial Science, was ecstatic about the team's success. "Various teams from our Department took part and the overall quality of the presentations was very good. I am looking forward for our Department to have an ongoing presence and involvement in the competition," he said.

The students walked away with a sum of R15 000 for winning the challenge.

Christiaan le Roux, David Rodwell and Robert Cronjé won the Hackathon data science challenge. With them is Mr August Carstens (far right), Data Science Manager at Capitec, one of the sponsors.

Financial Risk Management students write SAS certification exams

Carel van der Merwe

During the third term, Financial Risk Management students were given the opportunity to participate in the SAS Software Certified Young Professionals programme (SCYP). The SAS SCYP programme provides students free access to SAS software, online learning of SAS modules, preparation materials, and certification exams for four of their internationally accredited certifications. The four certifications include:

- SAS Certified Specialist: Base Programming Using SAS 9.4
- SAS Certified Professional: Advanced Programming Using SAS® 9.4
- SAS® Certified Statistical Business Analyst Using SAS® 9: Regression and Modelling, and
- SAS® Certified Specialist: Machine Learning Using SAS Viya 3.4

This was the first year that we included these exams as part of our modules, hence we had to employ a staggered approach to the offering. The final-year and honours students all wrote the Base Programming specialist certification, while the master's students wrote the Business Analyst certification. A mark of 725 out of 1000 is required to obtain the Base Programming certification, and a mark of 68% is required for the Business Analyst certification.

We are very proud of our master's class as all six of them obtained their Business Analyst certification with their first attempt. A further six students obtained the SAS base certification, with most of those that did not receive a high enough mark for the certification opting to rewrite at the beginning of next year. From next year, all four of the certifications will be integrated into our FRM programme. This will include the base certification as part of a final-year module, the advanced programming as part of the honours module, and the Business Analyst and machine-learning certifications as part of the master's programme. This would therefore allow a master's student to enter the workplace with four internationally recognised SAS certifications.

Additional to the SAS programming, the postgraduate FRM programme includes modules that focus on R and VBA in the honours year, as well as Python in the master's programme. Our students also have complete freedom as to which software they choose to do their research projects in. The Department would like to thank Mr André Zitzke who assisted with setting up the SCYP for our students.

More details on the programme can be found at: https://www.sas.com/sas/training/scyp.html

The FRM master's students are from left to right: Robert Cronjé, Delia Sandilands, David Rodwell, Christiaan le Roux, Nico Strauss and Kgomotso Sebitlo.

STUDENT ACHIEVEMENTS

Honours project presentations in 2019

Each year, the honours students have an opportunity to present their honours projects to the Department and other students. This year the presentations occurred during the week of 14 October. Below follows a summary of all the sessions as described by the various chairs:

Monday 14 October 2019
(Statistics and Mathematical Statistics)
Chair: Dr Chris Muller

This year's honours project presentations kicked off with a project investigating the marks of learners and a project investigating different pre-screening and sparse classification techniques.

In the first project, performance in Mathematics and Portuguese was predicted by classifying a student as either failing or passing the final period test. This was achieved through a comparative analysis of various classification techniques (logistic regression, decision trees, gradient boosting and support vector machines) to build a model with significant predictive power. Student results were also modelled using a linear regression model.

In the second project, different pre-screening and sparse classification techniques were applied in a high-dimensional data setting. Micro-array data, with the purpose of classifying cancer patients, were used to assess the different techniques. Analysis of variance and nearest shrunken centroid methods were the pre-screening methods considered in the project. Sparse and machine-learning classification techniques were compared, and the machine-learning models were found to outperform the sparse classification models.

Ulrich, Ishmat, Joslyn and Olivia

Tuesday 15 October 2019
(Statistics and Mathematical Statistics)
Chair: Prof Danie Uys

The first project presented reviewed the literature on quantile regression as an alternative and extension of least squares regression. It considered a quality of fit measure for quantile regression and investigated this measure on simulated data and on a practical data set. It compared the quantile regression coefficients to least squares coefficients. It was found that quantile regression produces a more in-depth understanding of the explanatory power of covariates.

The second project provided a review of the theory underlying the James-Stein estimator. The project further compared the James-Stein estimator with the sample mean vector if the population mean vector needs to be estimated. The comparison was done by applying the two estimators to simulated data as well as two real-life data sets.

The session was concluded with a project where the students applied machine-learning methods to the analysis of a Kaggle data set on student performance at two schools in Portugal. The objective was to predict or classify students at risk of failure in the final examinations of either Mathematics or Portuguese. Two machine-learning methods were chosen to compare, namely random forests and neural networks.

Aiden, Christiaan, Jared, Karla and Anchen.

Wednesday 16 October 2019
(Statistics and Mathematical Statistics)
Chair: Dr Surette Bierman

Two data science projects in collaboration with Capitec Bank were presented in this session. In the first project, the objective was to propose an accurate propensity model. The model should be able to accurately identify Capitec clients who are likely to make use of a particular savings product. In the second presentation, the aim was to propose a pre-delinquent model which could be used to accurately identify Capitec clients who are unlikely to meet their financial obligations in the future.

In both projects, the modelling process involved dealing with large data sets, the treatment of missing values and severe class imbalance, and a priori feature selection and dimension reduction steps. Several binary classification models were considered and empirically compared, based on their performance when used in combination with various options for each modelling aspect. The models ranged from linear classifiers such as LDA and L1 penalised logistic regression to non-linear models such as support vector machines, random forests and boosted models.

STUDENT ACHIEVEMENTS

Both groups of students succeeded in either reaching or exceeding the targets set by Capitec Bank. Positive feedback from the Capitec team ensured a rewarding experience, and we look forward to many more collaborative efforts.

PW, Ingo, Niël and Dylan.

Wednesday 16 October 2019
(Financial Risk Management)
Chair: Mr Carel van der Merwe

The first presentation looked at the importance of asset allocation and active management. A portfolio's total return can be decomposed into three components: (1) the market return, (2) the asset allocation policy returns in excess of the market return, and (3) the return from active portfolio management. This research project discussed the importance of asset allocation and active management and specifically delved into the performance attribution of multi-asset class portfolios. The purpose of this study is to dissect the returns of a portfolio and determine what aspect of the returns is attributed to the skill of the manager. The study found that a large percentage of returns from a multi-asset class portfolio come from strategic asset allocation.

The second project looked at the behaviour of GARCH model parameters under different market conditions using a rolling estimation window. In particular, it used the MIB30 Index, the S&P500 Index and Steinhoff stock price as data. These three series have different volatility clustering characteristics with some extreme events.

Thereafter a project that considered the determinants of credit spread movements was presented. In this project well-known methodologies in this field of research were applied to South African data.

The next group investigated how various rebalancing techniques can influence the risk profiles and profitability of balanced funds. Specifically, Masters' dynamic rebalancing was also tested among the traditional methods and mixed with various signalling methodologies. Even after trading costs, the dynamic methods proved to deliver optimal portfolios, and the best signalling method was the BEER (bond to equity earnings yield ratio).

Dynamic rebalancing techniques are therefore recommended for investigation by fund managers as possible replacements for the threshold and calendar rebalancing methodologies.

The fifth group looked at the behaviour of GARCH-model parameters when two extreme negative returns in close proximity were artificially added to a series simulated from a normal distribution with expected value zero and a given standard deviation. Dynamic estimation using a rolling estimation window was used and it was particularly interesting to see how the parameter estimates react when the extreme returns enter the estimation window and exit the window and also whether placing the extremes in the window has an effect on the estimates.

The final group investigated the investment behaviour of South African individuals through the use of probit models. It believed that investment behaviour has an important impact on the economy of the country. In order to improve the South African investment market, one needs to know to which extent certain factors influence investment participation. In their research, emphasis is placed on demographic and health factors. The amount of research done on this topic in a South African context is very limited, serving as motivation for this specific research. Through their research and model building, obvious factors such as employment and education are found to have profound positive impacts on investment. Interesting factors such medical aid, geography type and mobile phones also have a significant impact on investment decisions.

At the back are Shayna, Anton, Mandebe, Shannon, HP, Mangaka and Etienne. In front are Nathan, Simon and Guillaume.

Thursday 17 October 2019
(Statistics)
Chair: Dr Trudie Sandrock

During this session, six Statistics students presented their research.

The first project considered the use of machine-learning models to predict stock price movements. The objective of this study was to see if random forests, support vector machines and neural networks can be used to successfully predict the signals buy, hold and sell for a stock market index by using its historical data. Data from the S&P500 as well as the Nasdaq index was considered.

In the second project, multivariate time series were investigated by using the VARMAX procedure in SAS. A simulated VAR(2) model, the Lydia Pinkham Medicinal Company data set of Wei (2006) and the Terrorism data set of Enders (2015) were used to get a better understanding of the multivariate Box-Jenkins approach.

In the final presentation of the day, students presented results from a project considering a time series analysis conducted on windspeeds simulated from the WRF model in South Africa. The objective of this study was to identify the most appropriate SARIMA-type model for the data. In addition, the use of wind direction to improve forecasting accuracy of windspeeds was investigated.

James, André, Wiehan, Thamu, Nicola and Jan.

The Actuarial Science students also worked hard on their projects during the year, and Mr Stephen Burgess provides a brief overview below:

During 2019 the Actuarial Science staff and volunteers from industry supervised a number of Actuarial Science honours research projects. As far as possible, the students engage with issues that are topical in industry. This year, topics included an investigation of whether the increases in pensioners' cost of living is significantly different to the general Consumer Price Inflation, the impact of post-retirement investment strategies on income during retirement, and utilisation trends of medical members with diabetes. The projects provided our students with a challenge to apply their knowledge, built up over their four years at university, to a practical problem. Students also developed other important skills like problem solving, project management and communication.

Department delivers Top Magister Student in faculty

Top master's student 2019: Each year, Stellenbosch University (SU) gives recognition to students who performed exceptionally in academics, sport, service provision, leadership, culture and social impact. The award(s) received by a recipient is known as the Rector's Award for Excellent Achievement. At this event a medal for Top Magister Student is awarded to the top master's student in each faculty. At a special ceremony on 30 October, Annegret Muller received a Rector's Award for Excellent Achievement and the SU Medal for Top Magister Student (Economic and Management Sciences). We are extremely proud that one of our students received this award and the Department would like to congratulate Annegret with her outstanding achievements. Annegret is currently enrolled for a PhD in Mathematical Statistics.

MEET THE CLASS OF 2019

The Department of Statistics and Actuarial Science is proud of its third-year and postgraduate students.

These students have come a long way and we wish them well for the future.

Third-year students and lecturers

UNIVERSITEIT VAN STELLENBOSCH
DEPARTEMENT STATISTIEK EN AKTUARIËLE WETENSKAP
DERDEJAAR STUDENTE 2019

1st Ry: Dr S Bierman, Dr D Hofmeyr, Prof WJ Conradie, Dr H Viljoen, Me HM Cilliers, Mnr SJ Burgess, Mnr L Steyn, Prof PJ Mostert, Dr CJB Muller, Prof PG Slattery, Dr R Lötter, Dr MMC Lamont, Mnr R Clover, Dr MMC Lamont, Dr J Harvey, Dr T Sandrock
2de Ry: MM Perrott, S Phillips, FP Van der Walt, C Goncalves, YY Chen, HP Maletzkey, L Schreiber, M Kellerman, N Botha, L Dennis, S Van Graan, M Davids, PG Moehrke, MA Purchase, D Leibrandt, D Zembere, T Singh, J Narasimulu, A Kachikoti, J Bollman, A Korf
3de Ry: FJH Klomp, N Gihwala, N Malan, S Allschwang, M Du Toit, KH Van der Merwe, S Bijsters, E Frey, M De Beer, JN Stolk, A Eksteen, L Nel, GA Olbrich, MI Thomson, E Carinus, A Kotzé, M Carstensen, L Burger, L Bosman, C Van der Vyver, P Mkhatshwa, E Smit
4de Ry: J Rossouw, N Engelke, J Nieuwenhuizen, D Denner, N Le Roux, RA Tandau, K Campbell, VE Khumalo, J Möller, K Dennis, B Theron, M Boshoff, M Gerber, SJ Krause, GC Meyer, AJ Stofberg, N Breytenbach, C Weyers, M Dhanjee, Z Mwelase, H Van der Merwe, Dr F Kamper
5de Ry: J Bezuidenhout, Mnr CJ Van der Merwe, J Paul, K Endenburg, VA Pedlar, NS Mkhize, JD Vermaak, G Meyer, E Collins, SG Richardson, M Ramaisa, M Steyn, A Laubscher, Y Baijnath, JL Saunders, C Van Kamp, KSM Machiana, L Rall, B Rayner, A Joyner, Prof DW Uys, Prof SJ Steel

Photo credit: Anton Jordaan (SSFD/SCPS)

Postgraduate students and lecturers

UNIVERSITEIT VAN STELLENBOSCH
DEPARTEMENT STATISTIEK EN AKTUARIËLE WETENSKAP
NAGRAADSE STUDENTE 2019

1st Ry: Dr S Bierman, Prof WJ Conradie, Dr H Viljoen, Mnr C Van der Merwe, Prof T De Wet, Me HM Cilliers, Mnr SJ Burgess, Prof PJ Mostert, Dr CJB Muller, Prof PG Slattery, Mnr M Coxon, Dr R Lötter, Dr MMC Lamont, Dr J Harvey, Dr T Sandrock
2de Ry: CH Le Roux, J Ehlers, A Cronje, L Loock, A Bosch, K Laubscher, JG Greveling, O Heywood, N Taylor, D Vellema, N Leonard, H Giliomee, C Hattingh, P Hakutangwi, NM Prins, LC Loubser, MJ Bourdin, A Brand, S Roseveare, S Smit, M Ngiba, R Cronje, Mnr RJ Clover
3de Ry: Mnr D Corubolo, DT Rodwell, D Sandilands, J Malan, GJ Dreyer, AL Heyneke, ND Gounder, TC Lewis, I Ahmed, D Ascough, W Muller, WL Nel, J Joubert, AM Smit, JL Redelinghuys , AL Theron, LJ Schoonwinkel, M Smuts, A Muller, DC Van Zyl, TDP Mnyulwa, L Steyn
4de Ry: Dr D Hofmeyr, J Harley, KJ Sebitlo, W Van Zyl, C De Villiers, D Maxwell, I Von Gossler, J Swingler, U Kotze, EF Jacobs, S Venter, G Giliomee, HP Van Der Merwe, N Strauss, ZA Troester, L Nel, TC Lötter, J Pretorius, C Hendriks, BG Oosthuizen, K Bernhardt, M Giddings, C Thackwray, AJ Kirsten, MC Melonas, Prof DW Uys, Prof SJ Steel, Dr F Kamper

Photo credit: Anton Jordaan (SSFD/SCPS)

MEET AN ALUMNUS

Dr Wiesner Vos: Life as a data scientist at Google

Dr Wiesner Vos obtained a BSc degree at Stellenbosch University in 1997 – majoring in Mathematical Statistics and Actuarial Science – followed by BSc honours and master's degrees in Mathematical Statistics in 1998 and 2000. He completed a DPhil in Applied Statistics at the University of Oxford in 2005 as a Rhodes Scholar and he is affiliated with the Royal Statistical Society as a Chartered Statistician. Subsequently, he pursued a career in financial services and technology. He has been working as a data scientist at Google in London for the past eight years and is the co-founder and lead of an internal Google startup in their Area 120 programme, building cloud-based computational services for the investment banking industry.

Dr Wiesner Vos

Why did you choose Statistics?

Like many school leavers in South Africa at the time, I felt somewhat pressured to choose from a small number of highly regarded professions: actuary, medical doctor, chartered accountant, or lawyer. After a two-week stint as a medical student in 1995, I realised after a visit to Tygerberg Hospital that I was not cut out to be a doctor. Mathematics was my strongest subject at school, so I chose another one from the list, namely actuary. After a couple of semesters of Statistics, I was won over by the broad applications of Statistics and its intersection with so many other fascinating subjects.

I found Actuarial Science quite narrow, and the wider application of Statistics quickly peaked my interest. A subject that was often perceived as dull was brought to life by my lecturers at Stellenbosch University. Their enthusiasm for the subject and its many applications sparked my interest. We were in the early stages of modern Computational Statistics. I could hardly switch on a computer at the time, but I left the Department as a proficient R programmer and with deep statistical intuition, feeling equipped to put my knowledge to immediate practical application.

How did your career evolve?

I was fortunate to get an opportunity to read for a DPhil in Applied Statistics at Oxford when I left Stellenbosch, working with a leader in Computational Statistics, Prof Brian Ripley, who authored some of the textbooks we used at Stellenbosch. This is where my interest in solving large-scale applied statistical problems really developed strongly. I arrived at Oxford somewhat nervous, but I soon realised that my statistical education at Stellenbosch was second to none, and that I could easily hold my own against my peers from prestigious world universities thanks to that solid educational foundation.

After my studies and a short stint as a lecturer at Oxford, I returned to South Africa to work as an analyst at Woolworths Financial Services, and later the Incubeta Group (now NMPi). The latter opened a door to join Google in London as a data scientist.

For many years I worked as a data scientist in a research group in the Google Ads Engineering organisation that focuses on running large scale experiments and developing statistical technologies to measure the effectiveness of online advertising in Google Search and YouTube. During my later years in that group I rekindled my love for Statistical Computation, as I helped migrate our large team from R to Python and Tensorflow. R was simply not able to scale to the problems we wanted to solve and was hard to integrate in large-scale software engineering systems at Google. This was a steppingstone to another exciting opportunity at Google.

Google runs an internal incubator Area 120 that enables employees to build a small team and pitch a project idea to be funded as an internal startup. My brilliant Indian colleague, a former Goldman Sachs Quant, and I realised that Tensorflow is a great technology for scaling and accelerating derivatives pricing on Google Cloud Platform, that will help banks reduce the additional computational burden and cost that new Basel III and Basel IV regulations will impose on them.

Tensorflow is an open-source Google-developed computational library most typically used for machine learning, with a highly developed Python interface. We leverage it for its general computational benefits for creating Tensorflow-based derivatives pricing software that has recently been open-sourced on GitHub under https://github.com/google/tf-quant-finance. The performance of our tools has interested several large banks, with whom we are looking

to build partnerships to prove the technology at scale. Area 120 grants us significant freedom to build out our products and shape their development and has been a very exciting project to work on. The most challenging and rewarding aspects of my current role is to manage many complex relationships with internal and external stakeholders on our project, dealing with technical complexity and high levels of data security demanded by banks.

Dr Wiesner Vos on his Google bike

Is it all work and no play?

Most definitely not. Beside 5- and 9-year-old children demanding my time and attention, I enjoy sport very much. I completed the London Marathon in 2017 and play cricket for my local village cricket club. I also love music and try to attend a few music concerts or festivals every year. I am also very aware of the privileges I had growing up in South Africa, that so many South Africans do not have access to. I am very involved with an Irish-based charity Mellon Educate that build schools and educational programmes in the Cape Town townships. This has taken me to the heart of Khayelitsha several times as a construction worker to experience the hardship people face there. The charity is inspired by a Nelson Mandela quote: "Education is the most powerful weapon which you can use to change the world". This work has been very rewarding, and I have made many great friends for life.

What are some of your best Stellenbosch memories?

I fully embraced the social aspects of student life, sometimes a little too much perhaps. Despite that, I was always very dedicated to my studies. I remember several times that I forgot my practical work at home due to a combination of being scatter-brained and recovering from some student festivities the night before. This always involved me having to rush home on a hot sunny day by bicycle to collect (my always completed) work and returning in a pool of sweat to a wry smile by Prof Le Roux. I believe life is about balance, and still live by that motto.

How did Stellenbosch prepare you for your future career?

My lecturers at Stellenbosch encouraged me to take ownership of my work, and work through some nutty problems independently. They brought the subject to life with interesting practical applications, but always encouraged us to deeply understand the theoretical underpinnings of the work. This served me very well in the years that followed. Their passion for the subjects they taught was palpable, and they were always well prepared. I was very fortunate to be one of a handful of students in my postgraduate lectures, which was almost like an Oxbridge tutorial system. This gave me constant direct access to the professors who taught me, and they were never too busy to help when you went to their offices for academic help. I am also very grateful to my parents who encouraged me, and never pressured me. They allowed me to follow my own path and explore my passions.

What are the challenges and opportunities in Data Science?

Statistics is a vital component of Data Science and basic statistical principles still hold however much data you have to your disposal. Often data are generated by observational systems where petabytes of data do not necessarily help you in establishing causality. However, large amounts of data produced by modern technologies do open exciting opportunities for machine learning. Modern data scientists need to have a solid grounding in Statistics, Machine Learning and Computer Science. Strong programming skills and knowledge of modern technologies are vitally important to succeed as a modern data scientist. Modern statistical courses should have a strong computing component and give students exposure to R, Python, Julia, and lower level languages like C++. It is scary to think we are only on the cusp of the technological age as cars are about to drive themselves, medicine is about to be personalised, and computers are infiltrating every aspect of our daily lives. All of this generates data, the lifeblood of a data scientist.

The future for those considering a career in Data Science will be fascinating, challenging, and very rewarding.

UPCOMING EVENTS

Actuarial Science news

One of the highlights on the annual actuarial calendar is the Actuarial Society of South Africa's annual convention, which took place in Johannesburg on 22 and 23 October 2019. The range of topics presented at the convention continued to diversify into the wider fields of application of actuarial skills. Topics also covered new issues the profession needs to grapple with, such as the identification and pricing of cyber-risk and well-established but dynamic risks such as the outlook for longevity. Another important theme at the convention was the strong focus on addressing transformation in the South African actuarial profession. The Society also welcomed its new President, Mr Lusani Mulaudzi (a former student of the Department), who delivered an inspiring inaugural address.

Upcoming Seminars

SEMINAR PROGRAM: FIRST SEMESTER 2020	
Stellenbosch University Department of Statistics and Actuarial Science	
14 February	**Johané Nienkemper-Swanepoel** (Department of Genetics, SU) Visualisations for multivariate missing data
28 February	**Ingrid Van Keilegom** (Research Centre for Operations Research and Business Statistics, KU Leuven) On semiparametric modelling, estimation and inference for survival data subject to dependent censoring
13 March	**Zander Wessels** (NMRQL Research, Stellenbosch) Machine learning applications in investment management
3 April	**Adriaan van Niekerk** (Centre for Geographical Analysis, SU) The use of machine learning for making sense of satellite imagery
17 April	**Francois van der Bank** (Department of Industrial Psychology, SU) Using structural equation modelling to explain psychological mechanisms at work
8 May	**Kristian Muller-Nedebock** (Department of Physics, SU) Disorder in polymer networks

Seminars start at 13:00 and are held in room 2048 of the Van der Sterr Building, c/o Victoria and Bosman streets, Stellenbosch.

Enquiries: Prof Tertuis de Wet
Telephone: 021 808 3242
E-mail: tdewet@sun.ac.za

We hope you enjoyed this edition of the Department of Statistics and Actuarial Science's biannual newsletter. To access all our past newsletters and departmental information visit our website: http://www.sun.ac.za/statistics.

Follow us on social media: https://www.facebook.com/STATISTICSatSUN/

The next newsletter will be distributed in July 2020.