

Faculty Report on Postgraduate Activities in the Department of Statistical Sciences

The Department of Statistical Sciences has a large and diverse cohort of postgraduate students conducting research in topics spanning fields such as Statistics and Data Science, Advanced Analytics and Data Science, Applied Data Science, Operations Research, Ecological Statistics, Biostatistics, Epidemiology, Mathematical Statistics, Sports Statistics, Statistical Finance and many more.

During 2022, the honours cohort produced no fewer than 25 research projects conducted either individually or in pairs. Amongst these, two projects were selected as UCT's entries to the annual SASA honours project competition for 2022/3 for which we await adjudication. UCT students have an incredible track record in these competitions. Indeed, from the 2021/2022 cohort, Daniel Singer and Maxine Norrie won second prize for their project "Engineering Features for Tennis Match Outcome Prediction in a Statistical Learning Framework" under supervision of Mr. Stefan Britz. (UCT students placed 1st in the 2020/2021 running of the competition.)

We are also pleased to report a number of notable research activities from various Master's and PhD cohorts during 2022. For purposes of the present report, we outline some highlights with reference to relevant research groups in the department.

1. Postgraduates from the **Biostatistics Research Group**, headed by Prof. Francesca Little, observed another successful year in 2022 with PhD student Noelle Van Biljon winning the Young Statistician Showcase Award for presentation of PhD research on "Identification of Latent Classes of Anthropometric Growth and their Predictors within a South African Cohort" at the International Biometrics Conference, in Riga. Carlyle MCreedy was also given a Travel Award to attend conference and present his PhD research on "Early childhood wheezing phenotypes and determinants in a South African birth cohort study". A notable publication in 2022 again came from Carlyle with:

McCreedy C, Haider S, Little F, Nicol M, Workman L, Gray DM, Granell R, Stein DJ, Custovic A, Zar HJ. "The Early childhood wheezing phenotypes and determinants in a South African birth cohort: longitudinal analysis of the Drakenstein Child Health Study." Lancet Child & Adolescent Health 7 (2), 127-135, 2023.

For 2023, two of our brilliant young Master's candidates (and presently Assistant Lecturers in the department) will be attending the conference of the Royal Statistical Society in Harrogate, England. Heiletje Van Zyl will be presenting on "Analysis of Adapted Stepped-Wedge Cluster Randomized Trial for Parenting-Intervention Programme in Botswana", whilst Kelly Williams will be on "Investigating the influence of vaccine dose and prior Mycobacterium tuberculosis exposure on antigen-specific T cell responses using Generalized Linear Mixed Models."

2. The **Data Science Research Group/Cohort** encapsulates students from the Data Science Master's programme and interested parties. These students are typically working in industry and thus research often consists of industry-facing applied data science/statistics. Indeed, many of the students formally occupy positions as Data Scientists. Research is often focussed on research elements rather than sequences of peer-reviewed publications and cover real-world, real-time implementation of statistical technologies in the application of Data Science. During 2022, around 20 students graduated with dissertations covering a wide variety of research topics. Some of these research projects have been presented at national and international conferences, and some have been published in accredited journals. Notable international conference presentations include:

Baiju, V., Er, Ş., Dufourq, E. (2022). Word Sense Disambiguation in the Domain of Sentiment Analysis through Deep Learning. "International Conference on Data Science and Applications", Fethiye, Turkey, 7-11 September 2022.

Du Toit, C., Er, Ş., Salau, S. (2022). Cape Town road traffic accident analysis: Utilising supervised learning techniques and discussing their effectiveness. "4th International Conference on Advanced Research Methods and Analytics (CARMA)", 29 June – 1 July 2022, Valencia, Spain

The group was also represented by Alvin Reabow at an important local conference in Data Science, Deep Learning IndabaX South Africa:

Reabow, A., Er, Ş., Dufourq, E., Heany, S. (2022). Classification of Obsessive Compulsive Disorder using 3D Convolutional Neural Networks on structural brain MRI scans. "IndabaX conference at University of Pretoria", 12-15 December 2022.

Notable peer-reviewed journal publications include:

*CR Conrady, Ş Er, CG Attwood, LA Roberson, L de Vos (2022). "Automated detection and classification of southern African Roman seabream using mask R-CNN", **Ecological Informatics**, 69, 101593. <https://doi.org/10.1016/j.ecoinf.2022.101593>*

*Redelinghuys, J., Er, Ş. (2023). "soMLier: A South African Wine Recommender System (2022)", **Journal of Wine Research**, Vol. 34, Issue. 1. <https://doi.org/10.1080/09571264.2023.2184333>*

Finally, another notable initiative pertaining to the Data Science cohort concerns the STA-CHAI collaboration. In 2021, a relationship between Clinton Health Access Initiative, Inc. (CHAI) SA and the Statistical Sciences Department was formed to

create a Data Science fellowship programme which provides opportunities for students to work on policy-relevant questions in their dissertation year using government data. Specifically, the goal is to work directly with CHAI on a research project for the first 10 months of the dissertation year, thereafter submitting a report to CHAI before completing the dissertation. Present project abstracts can be found on the MASHA CHAI internships webpage:

<http://www.masha.uct.ac.za/masha/people/chaiinternshipstudents>.

3. The [Centre for Statistics in Ecology, the Environment and Conservation \(SEEC\)](#), headed by Prof. Res Altwegg, hosted the International Statistical Ecology Conference (ISEC2022) last year where PhD candidate Craig Mahlasi presented in a special plenary session on conservation in Africa and various postgraduate students presented research and co-chair sessions. Indeed, with the hybrid format of the conference, organisers decided to assign each session a co-chair which allowed us to promote our postgrad students in this role. SEEC also hosted their annual symposium which is aimed specifically at highlighting and promoting postgraduate research; All SEEC students present at these symposia and this year presentations were conducted successfully in person. A final notable mention for the SEEC group is the publication of Francois Becker's MSc in a very high-ranking journal:

Becker, F., J. Slingsby, J. Measey, K.A. Tolley, and R. Altwegg. 2022. Finding rare species and estimating the probability that all occupied sites have been found. Ecological Applications 32:e2505.

4. As expected, the [Modeling and Simulation Hub, Africa \(MASHA\)](#), one of the department's great success stories championed by A/Prof. Sheetal Silal, was very active where postgraduate activities in 2022 were concerned. Among other notable projects (see their research projects page [here](#)), MASHA launched the Malaria Modelling and Analytics: Leaders in Africa (MMALA) programme. The premise of the programme is to increase the number of PhD-trained mathematical modellers with malaria expertise in sub-Saharan Africa institutions, and foster relationships with NCMPs from Angola, Botswana, Eswatini, Mozambique, Namibia, South Africa, Zambia, Zimbabwe, Benin, Cameroon, and Ghana. As such, thirteen PhD and three Postdoctoral candidates were selected and have since taken up residence at MASHA where they will actively be developing research skills, take part in academic training, and conducting research. Notable publications from MASHA postgraduates include:

Awine T, Silal SP (2022) Assessing the effectiveness of malaria interventions at the regional level in Ghana using a mathematical modelling application. PLOS Glob Public Health 2(12): e0000474. <https://doi.org/10.1371/journal.pgph.0000474>

Patterson, J., Cleary, S., Silal, S.P. et al. A retrospective study assessing the clinical outcomes and costs of acute hepatitis A in Cape Town, South Africa. BMC Infect Dis 22, 45. (2022)

5. The **Statistical Finance Research Group**, under direction of A/Prof. Tim Gebbie, has produced a number of high quality research papers often backed by purpose-built software or general purpose applications. We invite interested parties to have a look at arXiv and ZivaHub (<https://zivahub.uct.ac.za/Statistical-Sciences>) for papers, software, and data sets. During 2022 statistical finance postgraduates published 6 new software and data research elements. We draw the attention to the arXiv author pages of Master's students [Patrick Chang](#) and [Ivan Jericevich](#) which contain multiple notable works. Patrick Chang has since moved on to a PhD programme at Oxford. A notable peer reviewed publication for the group is the publication of some of the MSc work of Ivan Jericevich in a high-ranking research elements journal:

Jericevich, I., Dharmesh, S., Gebbie, T. 2022, CoinTossX: An open-source low-latency high-throughput matching engine, SoftwareX, Volume 19, 101136

The rise to prominence of Statistics, Data Science, and Artificial Intelligence in industry and media during the last few years has certainly resulted in great demand for the technical proficiencies in statistics on which these fields are built. Consequently, the department has been successful in consistently graduating a large number of postgraduate students over the past few years. Of course the challenge of maintaining a high batting average in such a directly interdisciplinary domain is that this puts a significant supervision load on a relatively young academic contingent. This is both due to substantial competition from industry for individuals with these skills, where we find it difficult to recruit academics with the requisite qualifications to supervise and conduct research, as well as the contested nature of other interdisciplinary domains drawing directly from statistical theory and technologies.

A more practical challenge during 2022/3 has been that of loadshedding which has made both the coordination of research meetings and activities as well as physically conducting research more difficult for our students and academics. Apart from the reading and writing dimensions of statistical research, our work is very computational in nature for which extended periods (scheduled or otherwise) has the obvious detrimental effects.

Despite these challenges, we have been fortunate to have the opportunity to nurture great academic talent in our postgraduate students and share in their successes in a vibrant, caring, socially and environmentally aware active research department.