Information Required

Section A: Overview of the Research Project Proposal

1. Academic level of research project (Masters or Doctoral)
Doctoral

2. Broad field of research (Engineering or Astronomy/Astrophysics)
Astronomy

3. Title of the research project
Machine learning algorithm development for optimal follow-up of LSST transients with South African resources

4. Research project abstract/summary (max 250 words)

Both the SKA and LSST will unlock an exciting new discovery space for astrophysical transients. Early transient classification will be crucial for prioritising spectroscopic follow-up resources. Through our collaboration with the ANTARES (Arizona-NOAO Temporal Analysis and Response to Events System) transient broker team, we have the unique opportunity to develop our own algorithms optimised for finding anomalous targets ideal for follow-up by SALT, MeerKAT/SKA and other Southern African telescopes. This relies on access to LSST data and resources, not just the public alert stream. Our group at AIMS and the data science group at SARAO is heavily involved in the development of robust machine learning techniques for anomaly detection. Applying these to LSST data will put South African scientists and facilities in a position to take the lead in exploring and characterising newly discovered astrophysical phenomena.

Section B: Supervisor(s) Details

1. Primary supervisor's details

a. Title and full name
Prof. Bruce Bassett

b. Name of South African or SKA Partner Country university at which the primary supervisor is a permanent academic staff member
University of Cape Town/ African Institute for Mathematical Sciences/ South African Astronomical Observatory/ South African Radio Astronomy Observatory

c. Email address and/or contact telephone number (please note that in the event this project is approved, these contact details will be made available to students awarded SARAO postgraduate bursaries)
bruce@saao.ac.za

d. Supervision of postgraduate students – please provide the details of all the previous and current postgraduate students supervised. Please provide the information in table format, as shown below.

i. Doctoral Students

Name of student	Nationality	Date started Doctoral Degree (Month and Year)	Date completed / will complete Doctoral Degree (Month and Year)	Title of Research Project / Thesis	Co-Supervisor (if relevant)
Eli Kasai	Namibia	07/2015	12/2017		
Ethan Roberts	South African	03/2018	02/2020	Developments in machine learning and Bayesian inference for application in science	Nadeem Oozeer
Emmanuel Dufourq	Mauritian	05/2015	12/2018	Sentiment Analysis with Machine Learning	

ii. Masters Students

Name of student	Nationality	Date started Masters Degree (Month and Year)	Date completed / will complete Masters Degree (Month and Year)	Title of Research Project / Thesis	Co-Supervisor (if relevant)
Louwrens Labuscagne	South Africa	07/2017	12/2018	Investigate a collection of machine learning models to predict and provide insights into the Net Promoter Score (NPS) of a large Telecommunications company using network performance stats.	
Blake Cunningham	South Africa	07/2017	12/2018	Using an evolutionary algorithm approach to find optimal neural network-based policy approximation functions for reinforcement learning tasks	
Curtly Blows	South Africa	07/2017	12/2018	Reinforcement learning for Telescope Opimisation	
Zafiirah Hosenie	Mauritian	02/2016	07/2018		Nadeem Oozeer
Arrykrishna Mootoovaloo	Mauritian	07/2015	12/2017	Bayesian Inference for science with Meerkat and the SKA	
Kai Staats	USA	03/2014	07/2016	calable bio	

				regenerative model for human, off-world colonisation	
Lise du Buisson	South African	01/2013	07/2015	Machine Learning applied to Astronomy	

2. Co-supervisor / Research Supervisor's details (if relevant)

a. Title and full name
Dr. Michelle Lochner

b. Name of the university/institute, at which the co-supervisor/research supervisor is a permanent academic/research staff member
African Institute for Mathematical Sciences/ South African Radio Astronomy Observatory

c. Email address and/or contact telephone number (please note that in the event this project is approved, these contact details will be made available to students awarded SARAO postgraduate bursaries)
michelle@aims.ac.za/ 0741927641

d. Supervision of postgraduate students – please provide the details of all the postgraduate students supervised. Please provide the information in table format, as shown below.

i. Doctoral Students

Name of student	**Nationality**	**Date started Doctoral Degree (Month and Year)**	**Date completed / will complete Doctoral Degree (Month and Year)**	**Title of Research Project / Thesis**	**Co-Supervisor (if relevant)**

ii. Masters Students

Name of student	**Nationality**	**Date started Masters Degree (Month and Year)**	**Date completed / will complete Masters Degree (Month and Year)**	**Title of Research Project / Thesis**	**Co-Supervisor (if relevant)**
Kimeel Sooknunan	South African	January 2017	December 2018	Multiwavelength transient classification with machine learning	Prof. Bruce Bassett
Hope Moloko	South African	January 2018	December 2019	Bayesian source finding in optical and radio data	Prof. Bruce Bassett
Reem Elmahdi	Sudanese	August 2016	August 2017	Outlier detection with machine learning (AIMS research essay project)	

Section C: Full Research Project Proposal

Maximum of three A4 pages, written for a professional who is not necessarily an expert in the relevant subfield

1. Scientific merit: describe the objectives of the research project, placing them in the context of the current key questions and understanding of the field.
2. Feasibility: outline the methods that will be used to achieve the objectives. Provide details on the availability of required data / access to required equipment / availability of research facilities and other resources required. Include any relevant expected intermediate milestones and associated timeframes towards attaining the overall objectives of the project.
3. Link the proposed project to at least one SARAO research priority areas (refer to Annexure 1 of the Application Guide), and explain in some detail how the proposed research will contribute to the priority area(s).
4. If relevant, describe any particular qualifications, academic abilities, skills and/or experience that a student should have in order to successfully deliver on the objectives of the research proposed.

Research Proposal

Introduction

The Large Synoptic Survey Telescope (LSST) is expected to deliver 10 million alerts per night, triggered by transient events in the sky. While most of these will be artifacts, there will likely be thousands of potentially interesting astrophysical phenomena, including supernovae, AGN, variable stars and kilonovae. This presents an incredible technical challenge as these objects will require automated classification on a scale never before achieved in astronomy. We are fortunate that the field of machine learning has matured at the right time to take advantage of it for this golden era of transient astronomy.

There are already several possible “brokers” being developed for LSST. A broker will take in the alert stream from LSST, combine it with available information about the field in question, and apply a series of machine learning classifiers in order to filter the data stream. ANTARES (Arizona-NOAO Temporal Analysis and Response to Events System) is one of these brokers and already incorporates several sophisticated machine learning algorithms, including those developed by Lochner (see Narayan et al. 2017).

A key feature of LSST is that it is a photometric-only instrument: it takes images in multiple broadband filters, but lacks any spectroscopic ability that would usually be required to 100% accurately classify an object and learn more about it. This means that the most interesting or relevant astrophysical phenomena must be followed-up with other telescopes. The filtered alert stream from the broker will be available to anyone in the world to use, making it highly competitive to find the most interesting/ anomalous objects in the stream.

We have the unique opportunity in South Africa of having unprecedented access to exquisite multiwavelength follow-up facilities, including SALT and MeerKAT, as well as having access to LSST and connections to the ANTARES broker team (as well as other potential brokers). This means we will be able to tailor-make algorithms best suited for prioritising the types of astrophysical objects the South African community are most interested in and also which are best suited to follow-up.

Machine learning

The field of machine learning has experienced an explosion in recent years, matching the global explosion of available data, in both industry and science. Highly sophisticated algorithms have been developed that have definitively solved difficult problems such as image recognition and personalised recommendation engines.

Our data science groups at AIMS and SARAO specialise in machine learning and statistics, and have worked on multiple applications relevant for transient science (e.g. du Buisson et al. 2014, Lochner et al. 2016, Vafaei et al. 2018, Roberts, Bassett & Lochner in prep, Sooknunan et al. in prep). Part of this project will be further developing these algorithms and preparing them to work on incoming LSST data.

Putting South Africa at the forefront of astrophysical discovery

The first aim of this project is to essentially build the first astronomical recommendation engine, specifically tuned for South African capabilities and interests. The student will develop cutting-edge machine learning methods for transient classification and anomaly detection to produce each night a prioritised list of astrophysical phenomena available for follow-up. This involves incorporating user preferences (in this case the user is any astronomer wanting to make use of South African telescopes for follow-up) as well as which types of objects are most suitable for follow-up with what telescopes are available.

Due to Lochner's affiliation with LSST, we will have the opportunity to run these algorithms directly on LSST data (rather than the public, already filtered alert stream) as part of the transient broker, giving the South African community a competitive edge over other facilities.

Optimal scheduling with machine learning

The second component of this project is the development of the first "Intelligence Observatory". We will use machine learning to develop methods of optimally using all available facilities and resources for transient follow-up. This will become more important as more telescopes become roboticized that must make intelligent choices about observations. Treating all available facilities as a single AI system will result in a dramatic increase in efficiency over the traditional approach of using each independently with astronomers required to make the critical decisions. We have already seen with the exciting kilonova follow-up in 2017 that speed is critical in obtaining observations before the transient object fades.

The student will apply machine learning algorithms to this problem of optimising follow-up resources, which is also generally useful as an efficient approach to scheduling observations. The technique of reinforcement learning has already used to incredible effect to build an AI able to beat a world champion in the complex game of Go (https://deepmind.com/research/alphago/). Reinforcement learning is one of the very promising approaches we will explore in solving the telescope scheduling problem.

Feasibility

This project builds on extensive expertise within the AIMS/ SARAO data science groups. Both Bassett and Lochner are experts in machine learning, particularly as applied to transients. While obviously we won't have access to LSST data until it comes online, we will have access to preliminary test alert streams to test the approach, as well as simulations and early commissioning data. Of great use to us will be the Photometric LSST Astronomical Time-Series Classification Challenge (PLAsTiCC) data. PLAsTiCC is a community competition to develop algorithms to classify LSST transients, slated to start in September 2018 and conclude by January 2019, whereafter all the data will be made available. These will be a large set of simulations very close to what is expected from LSST.

Much of this project will require working with the existing facilities in South Africa. As Bassett is affiliated with SAAO, Bassett & Lochner are affiliated with SARAO, and Lochner is a member of ThunderKAT (the MeerKAT Large Science Project for transients), the student will have excellent access to all information and details required. The student will have access to office space at AIMS, SARAO and UCT and will join the data science groups at AIMS and SARAO, gaining exposure to experts in the field and a good collaborative environment.

Timeline

- March 2019 - The student will have completed relevant review of literature and familiarised themselves with machine learning.
- June 2019 - Through meeting with the South African transient science community, the student will gain an understanding of the classes of objects of particular interest, as well as clear knowledge of the capabilities of South African instruments.
- December 2019 - The student will gather a suite of existing machine learning algorithms commonly used for transient classification and applied them to several available datasets, including the PLAsTiCC data.
- June 2020 - The student will complete the framework for the South African prioritised transient recommendation engine and will publish this work.
- December 2020 - By this point the student should have completed preliminary work on general machine learning approaches to scheduling, including reinforcement learning.
- June 2021 - The student will finalise the machine learning scheduler and deliver it to the SAAO and SARAO communities
- December 2021 - The thesis will be written up

Student expectations and benefits

Any student applying for this project should have good python programming skills. Machine learning experience would be an advantage. This project will furnish the student with highly valuable skills, especially in machine learning, which are relevant in both academia and industry. The student would benefit from being part of the SARAO data science group, gaining excellent exposure to machine learning and statistics experts. The student will benefit from Lochner's connections and leadership position within LSST, allowing invaluable networking opportunities with USA-based scientists, as well as from Bassett's extensive international network of world-leading scientists. After completing this project, the student will be very well-placed as they will have the skills and experience required to work as a data scientist in industry or to work with South African astronomy facilities including MeerKAT and SALT or even to work in the USA or any LSST partner country on LSST data.

Section D: Signatures

1. Signature of the primary supervisor, with date

Prof. Bruce Bassett — Date: 31 August 2018

2. If relevant, signature of the co-supervisor/research supervisor, with date

Dr. Michelle Lochner — Date: 31 August 2018