Hawking Fellowship: Case for support

1 Track Record

I have achieved exceptional results at all stages of my education. At my highly competitive school I received the award for the best student, for achievement and engagement, in every subject I took at A-Level. I gained entry to the University of Bath where I studied Physics with a strong computational component gaining 1st Class Honours. I went on to do a Post Graduate Certificate of Education gaining valuable teaching and communication experience.

I began my PhD in Astrophysics in 2016 under the supervision of Dr Francesco Shankar. I began by working on a preliminary cosmological semi-empirical model, after gaining understanding of the field and modelling techniques I began to design and experiment with my own methodology. Current techniques have consistently struggled to faithfully reproduce the distributions and number densities of galaxies over multiple redshifts (Asquith et al., 2018). There are two major issues facing current models; firstly, over parameterization of models introducing moderate to severe degeneracies in the solutions (e.g. Lapi & Cavaliere, 2011). Secondly, these models work by simulating a discrete number of dark matter haloes which introduce a volume and resolution bias. The volume bias is due to the relative abundance of small to large haloes, for example to simulate ten massive haloes several thousand or even millions of smaller haloes must also be simulated. The resolution bias is due to the necessary cutoff in the smallest halo size due to the aforementioned high number density, in a large simulation the resolution must be limited to not exceed computational resource. These trade-offs mean it has been hard to simulate with good statistics of massive haloes without a low resolution cutoff to prevent simulating excessive numbers of small haloes. STEEL is a beyond the state-of-the-art technique, which I designed to better model satellite galaxies in groups and clusters by overcoming the issues suffered by traditional techniques. I list the innovations I have made with STEEL that are presented in Grylls et al. (2019) and my two upcoming papers (submitted and in prep).

- Developed a novel 'Statistical Dark Matter Accretion History' to remove volume and resolution bias from my simulations. Unconstrained by volume STEEL is able to model massive clusters missed by traditional models such as the remarkable high redshift cluster presented in Wang et al. (2016).
- Used the accretion histories and abundance matching to show to first order that dynamical time and not environmental processes effect distributions of satellite galaxies.
- Used the unique ability of a statistical semi-empirical model simulating the population averages to show that galaxy stellar mass estimates are inconstant with galaxy assembly models.

During my PhD I given talks in a variety of contexts, from a Lorentz centre workshop to the non-cosmology Euclid session at NAM 2017, I have also given an invited seminar to discuss STEEL at the University of Nottingham. I have co-supervised two summer students working on lenticular galaxies to create and implement empirical routines to model these elusive galaxies.

2 Research Proposal

Dark matter is ˜85% of the matter in the universe (Planck Collaboration et al., 2015); Leading theory states that the structure formation of dark matter is directly connected to the formation of galaxies. However, dark matter cannot be directly observed therefore understanding of the dark matter cosmological *paradigm is limited by the quality of observations that can be used as indirect measurements. It is expected that each dark matter halo forms a galaxy at its core. It is therefore possible to reveal the evolution of dark matter structure by tracing the formation and distribution galaxies across cosmic time. It is however unfortunate that major systematic errors in the analysis of galaxy observations undermine confidence in using galaxies as a tracer. The foremost systematic is in the estimation of galaxy stellar mass, I have shown using STEEL that current mass estimates are not consistent with dark matter assembly history. In this application I describe how I have highlighted and constrained this effect as well as the successes I have had in using STEEL to make self consistent predictions. This proposal shows how I can extend STEEL to ensure the self consistency of stellar mass estimates and dark matter assembly in future surveys. This work will be able to put constraints simultaneously on dark matter cosmological models, fitting models for data, and theories of galaxy assembly.*

2.1 State of the art and current limitations

The leading cosmological theory that describes the formation and distribution of galaxies is the ΛCDM cosmological model (see Bull et al., 2016, for a review). Notable successful predictions of ΛCDM cosmology include the Lyman-α forest, galaxy clustering and weak gravitational lensing. The traditional problems such as the 'cusp-core', missing satellites and 'too big to fail' have been mostly addressed by better simulations (coupling baryonic feedback at high resolution), better observations (observing faint satellites) or alternative theories (self interacting dark matter).

There are no current techniques to directly measure dark matter, observational tracers such as the distribution of galaxies are therefore key to our current understanding of dark matter structure. The current state-of-the-art method is to model the co-evolution of dark matter and baryonic matter (gas and stars that make up galaxies) to create galaxy mocks. The two leading techniques are Hydrodynamic simulations and Semi-Analytic models each uses a different method to include the dark matter component. In numerical simulations it is found that dark matter has a hierarchical evolution, from small perturbations in the initial density field large haloes grow under the influence of gravity smoothly accreting more dark matter as well as accreting other haloes in merger events. Hydrodynamical galaxy simulations co-evolve baryons directly with a dark matter simulation such that the gravitational influence of the baryons, as well as other feedback effects, can influence the dark matter. Semi-Analytic models use either the dark matter component from hydrodynamical simulations or an N-body dark matter only simulation to create merger trees[1], a simplification of the complex merging structure of dark matter.

In hydrodynamic simulations using a large number of cells the system is evolved by solving the equations of hydrodynamics, any process that would fall below the resolution of a given cell is then solved using sub-grid methods which are analytic routines that can be tuned to produce results that agree with observations. Hydrodynamical simulations are powerful tools and are able to give many physical insights into galaxy formation as they directly resolve structure formation and the output galaxies. However, the foremost limitation of such modelling is computational resource, each simulation can take many months to run, it is therefore required to make compromises in either the volume or resolution of the simulation. Furthermore, the sub-grid tuning can introduce degeneracies into the model obscuring the actual physical processes.

Semi-Analytic models initialise dark matter haloes with gas at high redshift then follow analytic routines to evolve the baryonic matter. The mergers of the galaxies are predominately dictated by the dark matter histories from dark matter merger trees. A major component of a Semi-Analytic model is the number of parameters that can be tuned to reproduce observations, less computationally expensive than Hydrodynamic simulations, Semi-Analytic models can be run many times to gain the best fit. However, as with hydrodynamical simulations such tuning can lead to degeneracies that obscure the actual physics.

[1] Alternatively analytic routines have been created to 'grow' merger trees that mimic simulations that grant flexibility in volume and/or cosmological parameters without the need to run multiple large volume simulations Parkinson et al. (2008)

Semi-Empirical models are a potential solution to the degeneracy introduced by the multi-parameter tuning inherent to the aforementioned techniques. Semi-Empirical models closely link observations to theory, in this instance ΛCDM dark matter merger trees, to create a model that by design reproduces specific observations. Commonly abundance matching, a technique used to predict the amount of stellar mass expected for a given halo mass, is used to populate dark matter haloes to reproduce the observed stellar mass function over multiple epochs. Additional physical assumptions are then gradually introduced adding the least assumptions/tuning parameters required to obtain a model consistent with observations. Building a model cautiously introducing additional assumptions/parameters limits the degeneracy from over parameterization improving the transparency of the model with regard to the important physics. Volume and resolution remain limitations of traditional Semi-Empirical models as they are built on the same dark matter merger trees as Semi-Analytic models.

My model STEEL is the first STastical sEmi-Empirical modeL. It builds on the low parameter modelling that make Semi-Empirical models powerful and further introduces a *statistical dark matter accretion history* removing the limitations inherent in traditional dark matter merger trees. STEEL has proven its potential with successes including accurate reproduction of the distribution of satellites in central haloes and the prediction of galaxy star formation rate. In addition to this the unique statistical nature of STEEL has allowed for the investigation of the self consistency of galaxy accretion in a hierarchical cosmological model. Removing discrete galaxies I explore the average satellite galaxy accretion histories and average central galaxy growth histories over the entire population. I find that many traditional galaxy mass estimates produce accretion rates that are inconsistent with central galaxy growth, I am therefore able to conclude that the hierarchical evolution of the dark matter and/or the galaxy mass estimates are flawed.

I propose STEEL be updated to become part of data processing pipelines to check derived products from observations in a given cosmology to ensure fitting model and cosmological model consistency. Once implemented this will be a **fundamental redesign to the connection between theoretical models and observational fitting in extra-galactic astrophysics**. Given the multitude of missions set to begin in the near future it is essential that models we use to fit data and theoretical models of galaxy formation are in agreement. Disagreement in these models is either a failure in the theory, observational fitting methods, and/or, an indication that ΛCDM assembly is incorrect. Results from STEEL will confirm or reject cosmology/fitting for surveys such as Euclid, an extra galactic mission with the main goal of providing precise cosmological constraints.

2.2 Motivation

The outcomes of this project are twofold: Firstly, STEEL will continue to be developed as a galaxy evolution model that will make testable predictions about galaxy populations derived from the dark matter accretion histories. Secondly, STEEL will be expanded in capability to be able to actively test the consistency of observational fitting models and ΛCDM cosmology. Through each of these pathways I will be able to place constraints on the ΛCDM cosmological model. Using STEEL I will also be able to test and develop theories of galactic formation and satellite evolution, with particular emphasis on the connection between mergers, active galactic nuclei, and quenching (the reduction of star formation rate). Finally, by using STEEL to investigate the systematics introduced by different cosmologies and observational fitting models I will be able to re-normalise data products from different surveys to correct for these effects.

2.2.1 Methodology of STEEL

STEEL can be broadly divided into four distinct aspects: Observational Data, Cosmology/Dark Matter, Galaxy Modelling, and, Outputs/Predictions. Figure 1 shows a cartoon of how steel works subdivided in these four categories. STEEL has successfully investigated satellite distributions, pair fractions and, star formation rates, however, for this cartoon we focus on the star formation rate as it best shows the

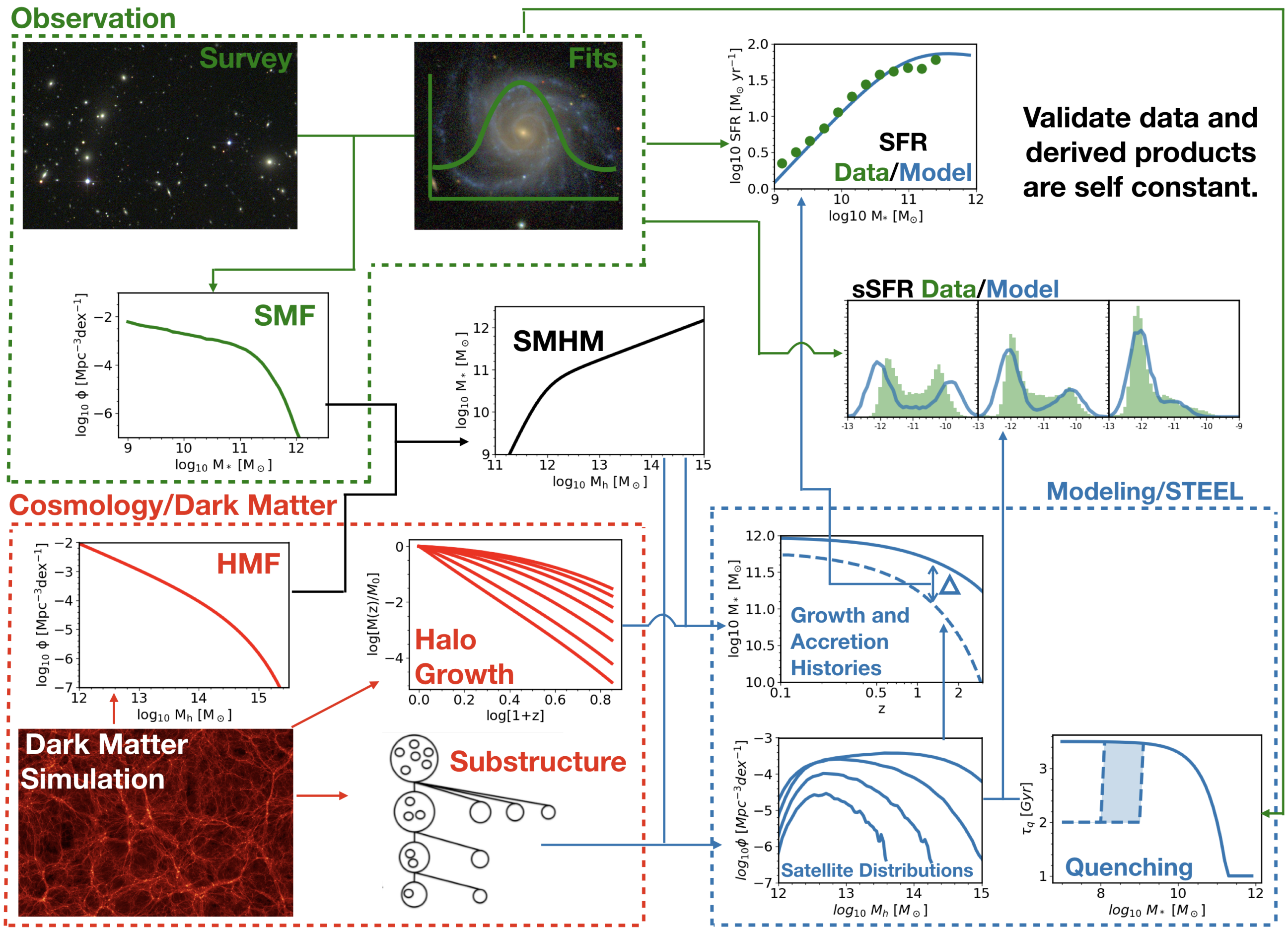

Figure 1: Schematic cartoon of how STEEL empirically models star formation rates. Described in full in Section 2.2.1.

interconnections of the empirical model. (In the following text colour directly corresponds to sections of the Figure 1.) By fitting and counting galaxies from surveys one can create the stellar mass function (SMF), the number density of galaxies as a function of mass. Using a halo finder to identify dark matter haloes a halo mass function (HMF), the number density of dark matter haloes as a function of mass, is extracted from dark matter simulations. Galaxies are assumed to reside in haloes with the same number density implying a mapping between galaxy stellar mass and host halo mass, this relation named the stellar-mass-halo-mass relation (SMHM) is the key empirical input used in STEEL . Using the stellar-mass-halo-mass relation and the dark matter substructure it is possible to derive the satellite distribution over multiple epochs. Using theoretical dynamical time arguments the satellite accretion history can be calculated, in addition the the central galaxy growth history is calculated using the stellar-mass-halo-mass relation and the halo growth histories. The difference between the the central galaxy growth history and the satellite accretion history can be attributed to galaxy growth via the star formation rate (SFR) which can in turn be compared to the observed star formation rate providing an additional and independent validation test to the model. Using an empirical quenching model derived from observations the full distributions of specific star formation rate (sSFR) can be modelled and compared to observations. This highly constrained and deliberate modelling is ideal for understanding the effects of individual theories in galactic astrophysics which would be very difficult to constrain via alternative, more traditional, approaches.

2.2.2 Using STEEL to check the self consistency of fitting models and cosmological models.

STEEL can be used to check self consistency of fitting models and observations. The follow steps provide an illustrative example for a consistency check using examples that I have had success with in my PhD:

- Observations are taken by a given mission e.g. Euclid, and a cosmology e.g. Planck is chosen to analyse the results in.
- Fitting models are used to derive galaxy properties over multiple epochs e.g. Stellar masses, Star formation rate, pair fractions...
- STEEL is run using a statistical dark matter accretion history calibrated on the chosen cosmology and the derived galaxy properties which are used to generate a SMHM relation and thus to assign galaxies of a given stellar mass/SFR to host dark matter haloes.
- STEEL then tests whether the evolution implied by the derived galaxy properties are consistent with the evolution expected from ΛCDM hierarchical assembly.
- There are multiple 'tiers' of results from STEEL: For example at its most basic the rate of satellite accretion can be compared to the rate of galaxy growth if this is not consistent (i.e there is more accretion than growth) it is reasonable to reject either the fits or ΛCDM cosmology. More complex modelling can predict star formation rates that would be required to maintain the galaxy growth and these compared to the observed star formation rate, should these disagree it is likely that our understanding of star formation processes or the fitting of star formation is incorrect.

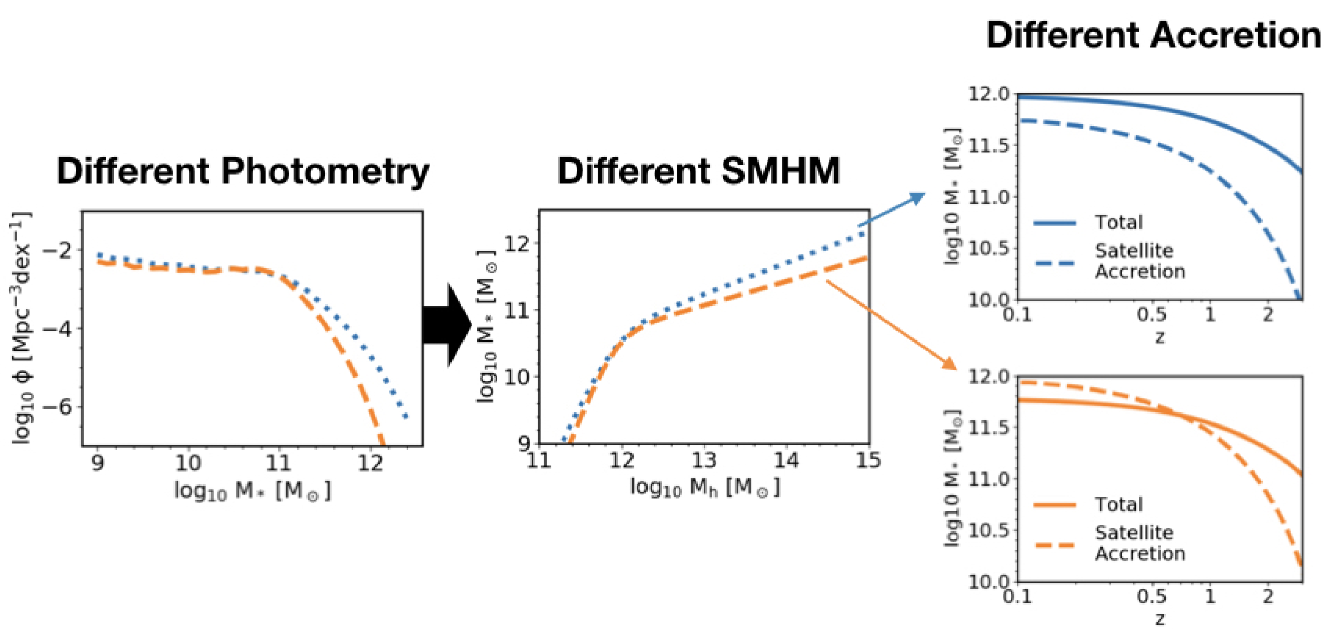

Figure 2: Schematic cartoon of how observational differences propagate into modelled inconsistencies.

Figure 2 is a simplification of the derivation of the satellite accretion vs. growth rate described in Section 2.2.1. In this example I show how a change in the observations caused by a photometric (fitting) choice propagates into an nonphysical satellite accretion that exceeds the central galaxy growth rate. This is an example of the lowest 'tier' analysis that can be used to check for consistency in data fitting models and ΛCDM assembly models, but it already rejects with good certainty traditional stellar mass fitting models (Grylls2019b, submitted). Advancing this technique to include more 'tiers' and tailoring it to additional galaxy properties of interest STEEL will bring theoretical understanding closer to the data we use to validate theory. This will simultaneously: increase the reliability of observations, constrain cosmological models, and advance understanding of galaxy assembly theory.

2.2.3 Galaxy modelling with STEEL

One of the major questions in the field of galactic astrophysics is how galaxies become quenched, i.e. what causes a galaxy to cease converting gas into stars. Several possible solutions have been discussed including AGN feedback, galaxy mergers, and/or, quenching due to inefficiency of gas supply though dark matter haloes above a given mass. Semi-Analytic and Hydrodynamical simulations obscure physics such as this due to the degeneracy of tuning parameters.

For example, if one were to attempt to test the connection between AGN feedback and star formation rate in a stable model without AGN it is likely that introduction of AGN will result in far more than just the re-tuning of starformation associated parameters. Due to the change in tuning it then renders the AGN and non-AGN models hard to compare fairly obscuring the connection between AGN and starformation. A Semi-Empirical model is better suited to solving these problems as they are capable of running with only the essential physics simply re-initialising essential galaxy properties to be consistent with observations at each time-step. In this way only the physical processes under investigation contribute to the results of the simulation making it extremely flexible, fast, and transparent.

An illustrative example of this technique is a simple model of galaxy growth in a central dark matter halo. At each time-step the central halo is assigned stellar mass using the stellar-mass-halo-mass relation. Introducing galaxy mergers by following the merging dark matter substructures the mass available to the central galaxy from accretion is calculated. If the accreted mass is sufficient to grow the galaxy to the size predicted by the stellar-mass-halo-mass-relation at later times then it is reasonable to argue that this galaxy grows mainly though mass accretion. If the accretion is too high then there must be more physics involved in the satellites i.e. stripping, or the merging model is wrong. If the accretion is too low to explain the growth of the galaxy then there is some additional mass growth mechanism i.e. star formation, or the merging model is wrong. With an empirical model understanding the physics and testing theories is much clearer.

Considering the above semi-empirical modelling example it becomes evident that higher-quality data will allow for more constraining power. For example, good star formation rate data could be used as a further input in the model to aid the growth of the galaxy, thus removing a degree of freedom. For an empirical model more data simply adds power, for a semi-analytic or hydrodynamical model it adds a constraint that if in disagreement with the model requires significant remodelling.

Open galaxy modelling questions that I will constrain during this fellowship are:

- Compare and contrast the effectiveness of proposed mechanisms for satellite galaxy quenching.
  - Central AGN feedback causing ram pressure stripping on satellites.
  - Stripping of the satellite gas reservoir by the halo environment.
- Understand the relative contributions of different processes to central galaxy quenching.
  - Mergers causing morphological change associated with the end of starformation.
  - AGN feedback suppressing star formation directly or coupling with the halo environment.
- Distinguish which populations of galaxies are transformed into lenticulars and by which mechanism(s).

With each of these investigations the modelling process is to take theoretical models, often interpreted from observations or small high resolution hydrodynamic simulations, and apply it to a statistical model to understand how these effects would manifest in the full galaxy population.

2.3 Impact of Research

In addition to being a novel approach, STEEL has also been devolved with several upcoming large surveys in mind. EUCLID, LSST, eROSITA, Athena, and JWST each have the capability to greatly impact our theoretical understanding of galaxy formation and cosmology. However, as shown with my previous results without accurate theoretical modelling to inform the data fitting models, these surveys risk the same inconsistencies as previous missions, sacrificing potential impact at great fiscal cost to the scientific community. In addition, fitting/cosmologies have previously been corrected with overly simplistic factors. For example to convert from a Chabrier (Chabrier, 2003) initial mass function to a Salpeter (Salpeter, 1955) inital mass function multiplicative factor of 0.63 is used and cosmological corrections use only a factor called 'little h' (Croton, 2013). I have shown how the propagation of inputs create a requirement for more subtle corrections in shape and normalisation. STEEL can be used to create cosmology/fit corrections that capture these subtleties and allow years of old data to be used with the high comparative quality that STEEL will provide to new surveys. Current models struggle to produce the volumes currently observed so are unlikely to keep up with the larger volumes soon to be observed where as for a statistical model this is a non-issue. STEEL is ideally and uniquely suited for the future of galactic modelling and the advancement of understanding the formation and nature of our Universe.

2.4 Research Plan

For this project to have maximum impact it will require strong early collaboration with groups such as the Euclid consortium. STEEL is already known by several influential empirical modellers. My research group in Southampton is heavily involved in the Euclid collaboration for which STEEL will represent an invaluable tool to produce reliable "mock" test galaxy catalogues and to properly interpret the data. This is an imminent vital task given that Euclid is currently scheduled for 2022.

The first goal is to use STEEL to make a tool that checks fitting model and cosmological model consistency. Working with contacts from the Euclid collaboration this can be prepared with the flagship mocks such that the tool is ready for launch. Following launch we will have access to consistency checked surveys which will greatly increase the empirical power of the model. With this highly constrained tests of galaxy formation theories can be carried out with STEEL to place constraints on the quenching of galaxies and other open questions.

To ensure maximum impact via collaboration and awareness during the first year I will organise a workshop including modellers and observers. The aims of this workshop will be:

- Publicise the 'fitting model'-'cosmological model' inconsistencies. Make observers aware of how this should influence the way they approach fitting. Make modellers aware of how the data should be considered when comparing to models.
- Gain support from those people working on large surveys to adopt STEEL as a method to help develop consistent fits.
- Build discourse between observers and theorists about the open problems and the systematics present in both observation and modelling that are barriers to making clear theoretical progress.

3 Professional Development

The University of Southampton provides many courses to ensure the continued professional development of their staff. Attending the following courses during the first half of the fellowship will enhance my ability to produce and communicate my science.

- Optimising your time: With an ambitious project such as this it will be important to me to improve my already strong self-management skills to ensure timely progress.

- Successful communication: I will continue to improve my communication skills to ensure an efficient transfer of knowledge of the core modelling concepts behind STEEL create maximum impact.
- Peer Review: The process of peer review is essential to develop outstanding skills in reading and writing high-quality papers.
- Quality Papers: Papers are the foundation of a good career and are the core material by which work is judged and recognised. With continual development I will further improve my papers to ensure my techniques are visible and accessible to the community.
- Writing Science for the public: The scientific community are funded by the public and have a duty to report their findings in a timely and engaging way. This course will allow me to ensure my research is well received. To provide continual relatable content I will make a monthly blog on what I am working on pitched toward general public.

In addition to these courses I also have strong links to the Research Software Engineering community keeping me up to date with the latest in best computational practice. My models and codes will be efficient as possible but also well documented, this means my work is accessible to the modelling community.

4 Public Engagement

4.1 Track Record

I have a wealth of experience in public engagement. Since starting University I have provided over 100 hours of engagement activities which include:

- Bath TAP: General science focusing on forces. Age range 5-12 years.
- BRLSI: The 'Bright Spark' workshops focusing on electromagnetism. Age range 7-14 years.
- Planetrella: Talk and a live demo of how plasma is contained by magnetic fields used to explain the relevance of the Aurora to our daily lives and the history of their discovery. Age range 4-90 years.

In addition to this I have a teaching qualification in Mathematics for the 11-18 age range so I am comfortable disseminating ideas to young audiences. In addition to my outreach work I have also spent over 50 hours volunteering in a school for autistic children so have familiarity working with less able people.

4.2 Plans and experience engaging audiences

4.3 Activities

STEEL is a lightweight model that can be configured to run very quickly. We can use this to create demos where people interact with the actual scientific code. The following are examples of demos that I expect to develop to be able to to engage the general public with the beauty of galaxy modelling.

- Use STEEL combined with the outreach tool ASTERA, an interactive cosmological 3D galaxy visualisation program, to create image mocks of galaxies in the Universe at any epoch given different stellar-mass-halo-mass relations. Asking the public to change the stellar-mass-halo-mass parameter they can try and match the look of real clusters or create 'extreme universes'.
- One of the joys of empirical modelling is that some of the key results can be approximated very simply. With the rise in schools teaching basic programming this should make GCSE and A-Level students be able to code and access the results with minimal extra intervention. By providing simplified halo and galaxy catalogues it will be possible to create a workshop to teach the basic principles of empirical modelling.

4.4 Training

In addition to the professional development the University of Southampton offer outreach specific courses. Before beginning the public engagement activities I will take these three courses to ensure I am up to date with effective engagement strategies.

- Introduction to Public Engagement with Research
- Engagement with schools
- Introduction to one to many engagement.

References

Asquith R., et al., 2018, http://dx.doi.org/10.1093/mnras/sty1870 Monthly Notices of the Royal Astronomical Society, 480, 1197

Bull P., et al., 2016, Beyond ΛCDM: Problems, solutions, and the road ahead, http://dx.doi.org/10.1016/j.dark.2016.02.001 doi:10.1016/j.dark.2016.02.001, http://arxiv.org/abs/1502.01589 http://dx.doi.org/10.1051/0004-6361/201525830 http://www.aanda.org/10.1051/0004-6361/201525830

Chabrier G., 2003, http://dx.doi.org/10.1086/376392 Publications of the Astronomical Society of the Pacific, 115, 763

Croton D. J., 2013, Publications of the Astronomical Society of Australia

Grylls P. J., Shankar F., Zanisi L., Bernardi M., 2019, http://dx.doi.org/10.1093/mnras/sty3281 MNRAS, 483, 2506

Lapi A., Cavaliere A., 2011, http://dx.doi.org/10.1155/2011/903429 Advances in Astronomy, 2011, 1

Parkinson H., Cole S., Helly J., 2008, http://dx.doi.org/10.1111/j.1365-2966.2007.12517.x Monthly Notices of the Royal Astronomical Society, 383, 557

Planck Collaboration et al., 2015, http://dx.doi.org/10.1051/0004-6361/201525830 Astronomy & Astrophysics, 594

Salpeter E. E., 1955, http://dx.doi.org/10.1086/145971 The Astrophysical Journal, 121, 161

Wang T., et al., 2016, http://dx.doi.org/10.3847/0004-637X/828/1/56 The Astrophysical Journal, 828