Dr. David Jones
Center for Astrophysics on La Palma
Insituto de Astrofísica de Canarias
E-38712, Santa Cruz de Tenerife, Spain
Telephone: (+34) 922 605 200 Ext. 3718
Email: david.jones@iac.es

9 September 2024

To whom it may concern:

It is my pleasure to provide the following report on Mr. Milan Pešta's doctoral thesis.

The thesis focusses on the application of state-of-the-art machine learning methods in understanding binary stars through photometric observations – a very timely topic given the current abundance of time-resolved ground- (ASAS-SN, OGLE, SuperWASP, ZTF, etc.) and space-based (Kepler, TESS, Gaia) photometric surveys as well as the impending arrival of the Vera C. Rubin Observatory and its Legacy Survey of Space and Time. With such high quantities of data, it is impossible to envisage large-scale analyses without the use of machine learning/artificial intelligence. Similarly, the thesis takes advantage of recent advances in the fidelity of numerical models of binary stars (namely the `phoebe2` software package, first released in 2016) as a basis to relate the observed photometric variability to the physical properties of the systems analysed.

Beyond the obviously timely nature of the work presented in this thesis, it is clearly of a high standard with the results of two chapters already being published in peer-reviewed journals – one with Mr. Pešta as first-author in Astronomy & Astrophysics – and the results of a third chapter submitted for publication as first-author also in Astronomy & Astrophysics (which I have no doubt will be accepted for publication in due course). The thesis is well-written and clearly structured, demonstrating a clear understanding of the relevant literature as well as the implications of the results. The thesis represents a strong *independent* contribution to the field, making Mr. Pešta deserving of the opportunity to defend.

Below I add a few more specific comments and questions, which I would like to present to Mr. Pešta during his defence:

Chapter 1 introduces the Roche model for binary stars and the different classifications of binary based on orbital separation and light curve shape - with particular emphasis on contact binaries and dark companion binaries. The chapter includes significant historical background, demonstrating the author's familiarity with the literature. In section 1.4, it is claimed that it is highly cost-inefficient to obtain astrometric measurements for large numbers of dark binary candidates without pre-selection of only the most promising candidates – is this true in the *Gaia* era?

Chapter 2 presents the primary methods of data analysis employed in the thesis. `phoebe` is used for the synthesis of model light curves which can then be used to infer, using Bayesian statistics, the population parameters of real contact binaries based on their observed photometric amplitudes (presented in Chapter 4). Similarly, random forest classification is outlined, which is later trained using a set of synthetic light curves to investigate the possibility of identifying dark companion binaries using photometric survey data. My fundamental question regards the construction of the underlying models. Contact and heavily tidally distorted binaries are notoriously difficult to model (see e.g. Kochoska et al. 2019). In `phoebe` terms, modellers often run into issues with the surface properties of the stars leaving the bounds of the input stellar atmosphere models. This issue is briefly mentioned in Chapter 5, but I wonder more generally was this an issue also for the work in chapter 4? Was any fine tuning required/performed?

Chapter 3 investigates the consequences of parameterisation choices on binary light curve solutions using the `phoebe` software package and comparing models of AI Phe with those derived in Maxted et al. (2020). The primary conclusion is that, while parameterisation choices do indeed matter, the principle source of discrepancy between the results of Maxted et al. and those presented here are due to the mapping of `phoebe` native parameters to the parameters employed by the chosen backend (`ellc`). The differences between the native parameters are not discussed nor is the methodology for the mapping – it would be interesting to understand this and how this motivates the conclusion that the `phoebe` parameterisation should be favoured. Fundamentally, the results indicate that particular `phoebe` parameters *could* be mapped to different `ellc` parameters, given that run A of Maxted et al. uses only `ellc` and apparently similar parameters in `phoebe` mapped to the `ellc` backend lead to different results.

Chapter 4 applies Bayesian procedure to constrain the mass ratio distribution of a sample of contact binaries selected from the Kepler Eclipsing Binary Catalog via a Period-Luminosity-Colour relationship. As mentioned in my comment re: Chapter 2, modelling of compact binaries is notoriously difficult. The `phoebe` model is initialised with equal temperatures (approximately solar) and default limb-darkening and gravity-brightening coefficients. The choice of atmosphere model is not mentioned, but presumably is the default Castelli & Kurucz option. Do these choices constrain the parameter space which could be explored (i.e. are certain regions of parameter space excluded as models fail in this region?). Similarly, the choice of atmosphere and temperature is more representative of late-type stars than early-type which will have significantly larger temperatures and atmospheres which may not be well represented by the Castelli & Kurucz model. Although it is clear the use of a distinct PLC will help to mitigate some of these issues, how might the choice of parameters be expected to skew the results?

Chapter 5 investigates the potential of machine learning techniques, trained by synthetic light curves of dark companion binaries and dark companion impostors, to identify dark companion binaries among other ellipsoidally modulated variables. Here, the authors do mention that due to the very broad parameter space, some models do not produce a valid light curve. I suspect that these occur on the edge cases, near to transitions between non-Roche-lobe-filling configurations, semidetached and contact binaries. This could potentially introduce discrete jumps in the parameter space covered, perhaps leading to an artificially increased efficiency of identifying dark companion binaries.

The final chapter succinctly summarises the results of the thesis and highlights well Mr. Pešta's contribution to the field. Ultimately, I would like to congratulate Mr. Pešta for the work presented here and strongly recommend he be allowed to proceed to the defence of his thesis.

Regards,

Dr. David Jones
Instituto de Astrofísica de Canarias