Dr hab. Agnieszka Kuźmicz

Astronomical Observatory

Jagiellonian University

Kraków, 11.06.2026

JAGIELLONIAN UNIVERSITY IN KRAKOW

Faculty of Physics, Astronomy and Applied Computer Science

Review of the doctoral dissertation of Msc. Anjitha John William

Entitled:

“Photometric redshifts and clustering statistics of the large-scale structure”

The doctoral dissertation of Ms Anjitha John William presents her original research on the large-scale structure of the Universe using photometric surveys of galaxies and quasars. Its main focus is the development of advanced machine learning methods for estimating photometric redshifts, which are essential for cosmological analyses when spectroscopic measurements are unavailable. The thesis introduces the Hybrid-z machine learning framework, combining multi-band photometric data to improve photometric redshift accuracy in the KiDS-DR4 dataset. Using these improved redshift estimates, Ms William analyses the clustering of galaxies and quasars to study their connection to the underlying dark matter distribution. The thesis also examines non-Gaussian properties of matter clustering and the dependence of clustering strength on galaxy properties such as stellar mass and colour.

The submitted doctoral thesis consists of 6 chapters, of which chapters 3 and 4 are reprints of papers already published and under review in the *Astronomy and Astrophysics Journal.* The contribution of Ms William to the preparation of the articles is significant according to the authors' declaration.

Chapter 1 introduces the theoretical and methodological background of the dissertation. It reviews gravitational theory in general relativity for an expanding Universe and outlines key observational concepts such as redshift and the Hubble–Lemaître law. It also presents structure formation and matter perturbations, and introduces statistical tools for large-scale structure analysis, including correlation functions and counts-in-cells methods.

Chapter 2 presents the basics of photometric redshift estimation. It compares spectroscopic and photometric redshift methods and reviews template-based, empirical, and machine-learning approaches. Particular attention is given to neural networks, especially artificial and convolutional neural networks, applied to photometric data and galaxy images.

Both chapters are comprehensive, coherent, and clearly structured, however, the level of detail is not fully consistent across the text. In some places, it is very detailed, while other parts rely on assumptions without sufficient clarification. There are also minor issues with logical structure: some quantities are introduced before being fully defined or properly motivated (e.g. the Hubble distance in Eq. 1.26 or abbreviations such as

ul. prof. Stanisława Łojasiewicza 11
PL 30-348 Kraków
tel. +48(12) 664-48-90
fax +48(12) 664-49-05
e-mail:
wydzial.fais@uj.edu.pl

SED). The logical flow is sometimes unclear (e.g. on page 7, where the text suggests that the definition of redshift follows from $f=c/\lambda$, whereas redshift is in fact defined observationally by spectral line shifts).

Chapter 3 includes an already published paper in which Ms William presents a deep machine learning-based approach to photometric redshift estimation for the KiDS-DR4 bright galaxy sample. It introduces a hybrid model, *Hybrid-z*, which combines a convolutional neural network applied to galaxy images with an artificial neural network applied to multi-band photometric magnitudes. The method achieves a substantial improvement over ANNz2 framework, reducing the photo-z scatter by ~20% to ~0.014(1+z), while keeping very low bias ($\sim10^{-4}$) and demonstrating robust performance across multiple independent spectroscopic datasets. A key contribution is smoothing the GAMA redshift distribution, reducing training-set artefacts, and improving the reliability of the photo-z estimates. The published paper shows a strong understanding of current trends in machine learning for astrophysical applications, including the limitations and potential systematic effects associated with training sets and deep learning models. The author also demonstrates motivation for further development and extension of the proposed methodology, including its application to multi-band imaging and large-scale surveys such as KiDS-DR5 and LSST.

In Chapter 4, the author presents a paper submitted to Astronomy & Astrophysics, which is currently under review. The study investigates clustering properties across a wide redshift range up to z = 2.7, using the author's *Hybrid-z* framework to estimate photometric redshifts for quasars from the KiDS-DR4 catalogue. Based on this analysis, the author measures the evolution of quasar bias, demonstrating that the analysed quasars reside in dark matter halos with typical masses of approximately $10^{12}\ h^{-1}\ M_{\odot}$. The study also discusses systematic effects which influence the accuracy of bias measurements, emphasizing the importance of accurate redshift calibration. An additional outcome of the study is the catalogue containing 157,000 quasars. Overall, the results of the paper are consistent with previous studies, reporting a comparable redshift evolution of quasar bias and typical dark matter halo masses. Concerning this section, I have a few comments and questions regarding the analysis presented in the paper:

- In the Introduction, author writes that "QSO spectra are dominated by non-thermal emission, giving them colors distinct from other extragalactic sources" – this statement seems unclear in the optical regime, where emission is typically dominated by a quasi-thermal accretion disc component.

- Stellar contamination is corrected using the angular 2PCF of "candidate stars" from Nakoneczny et al. (2021). Could the author clarify why a candidate star sample was used instead of, e.g., a known/confirmed star sample, and what fraction of the candidate star sample may be misclassified?

- The uncertainties in the quasar bias shown in Fig. 6 appear to be systematically smaller in other studies compared to those obtained by the author. Could the author comment on the origin of these differences?

- How strongly do the estimated quasar bias results depend on the assumed cosmological model?

- To what extent, the bias evolution and halo mass estimates are driven by the assumed redshift distribution rather than the clustering itself?

Section 6 presents unpublished results on higher-order clustering statistics of galaxies in KiDS-DR4 using the Count-in-Cells methodology. Based on the KiDS-DR4 bright galaxy sample and photometric redshifts, Ms William investigates non-Gaussian features of the galaxy distribution and their dependence on redshift, galaxy colour, and stellar mass. The results show that red galaxies and high-stellar-mass galaxies exhibit the strongest higher-order clustering and non-Gaussian features, consistent with their presence in more massive dark matter halos. The study also demonstrates that the galaxy distribution becomes increasingly Gaussian on larger angular scales. Additionally, an upturn in the clustering statistics at large scales is detected, which is likely driven by observational systematics and survey geometry effects.

Regarding this chapter, I have some editorial comments, particularly concerning the order in which the figures are placed. Several figures appear before they are referenced in the text (Figures 5.2, 5.3, 5.4, 5.5, and Figure 5.6, which is placed in an earlier subsection). Additionally, Figure 5.9 should be placed immediately after Figure 5.6. In Section 5.3.5, the ordering of the figures is also inconsistent — Figure 5.11 should be numbered as Figure 5.13 by analogy with the previous subsection: first redshift bins, then galaxy color, and finally mass bins.

I also have some minor comments regarding this chapter. Ms William uses the u-g vs. M_r diagram to separate blue and red galaxies without explaining how the absolute magnitudes were determined (i.e., how the galaxy distances were estimated). Also, unfortunately, Ms William did not consider easy-to-implement extensions of the analysis, which she mentioned in Section 5.5 ("Future directions"). For example, the analysis could be extended by dividing the red, green, and blue galaxy populations into stellar mass bins. Such an analysis would not require substantial additional effort. In addition, using mock catalogues to validate the CiC measurements and quantify the impact of survey geometry and masking could significantly improve the conclusions about the higher-order clustering measurements.

The dissertation ends with a summary section.

Summing up, the doctoral thesis "Photometric redshifts and clustering statistics of the large-scale structure" by Anjitha John William makes valuable contributions to the field of observational cosmology and machine learning methods. The thesis presents a coherent structure, ranging from methodology to application. The clustering statistics were applied to the KiDS survey for the first time, and an additional outcome of the thesis is the publicly available Hybrid-z framework, as well as a KiDS quasar catalogue.

In my opinion, the doctoral thesis of Ms Anjitha John William meets all formal and customary requirements for a doctoral thesis. Therefore, I recommend that Ms William be admitted to the next stages of the doctoral procedure, including the public defence.

Agnieszka Kuźmicz

JAGIELLONIAN UNIVERSITY IN KRAKOW

Faculty of Physics, Astronomy and Applied Computer Science

ul. prof. Stanisława Łojasiewicza 11
PL 30-348 Kraków
tel. +48(12) 664-48-90
fax +48(12) 664-49-05
e-mail: wydzial.fais@uj.edu.pl

Center for Theoretical Physics
Polish Academy of Sciences
Aleja Lotników 32/46, 02-668 Warsaw
Tel. (+48 22) 847 09 20, Fax/Tel: (+48 22) 843 13 69
E-mail: cft@cft.edu.pl, NIP: 525-000-92-81, REGON: 000844815

Konkluzja recenzji rozprawy doktorskiej
(Conclusion of dissertation review)

„Photometric redshifts and clustering statistics of the large-scale structure"

Tytuł rozprawy *(Dissertation title)*: ……………………………………………………………………………………………………..

Anjitha John William

Autor rozprawy *(Author of the dissertation)*: ……………………………………………………………………………………………

☑ Pozytywna ocena *(Positive conclusion)*:

Stwierdzam, że przedstawiona mi do recenzji rozprawa spełnia wszystkie wymagania ustawowe i zwyczajowe stawiane rozprawom doktorskim i wnoszę o dopuszczenie jej do dalszych etapów postępowania doktorskiego, uwzględniając publiczną obronę.

*(I conclude that the presented dissertation meets the formal and customary requirements for doctoral dissertations and I recommend its admission to subsequent stages of the procedure, including the public defense.)**

☐ Ocena negatywna *(negative conclusion)*

Stwierdzam, że przedstawiona mi do recenzji rozprawa nie spełnia wszystkich wymagań ustawowych i zwyczajowych stawianych rozprawom doktorskim i dlatego nie rekomenduję dopuszczenia jej do dalszych etapów postępowania doktorskiego.

*(I conclude that the presented dissertation does not meet the formal and customary requirements for doctoral dissertations and therefore I do not recommend its admission to subsequent stages of the doctoral procedure.)**

Uzasadnienie powyższej oceny znajduje się w raporcie będącym załącznikiem 1.
(The justification of the above assessment can be found in the detailed report in the attachment 1.)

11.06.2026 Agnieszka Kuźmicz
…………………………..
Data i podpis
(Date and signature)

Załącznik 1: Recenzja rozprawy doktorskiej
(Attachment 1: Review of the dissertation)

**Zaznacz ocenę (Please tick the box with your conclusion)*