Review of Anjitha John William PhD Thesis "Photometric Redshifts and Clustering Statistics of the Large-Scale Structure"

The doctoral dissertation of Anjitha John William addresses one of the fundamental challenges of modern observational cosmology: the extraction of reliable redshift information from large photometric surveys and the use of these data for studies of the large-scale structure of the Universe. The work combines modern machine-learning techniques with observational cosmology and applies them to the Kilo-Degree Survey Data Release 4 (KiDS-DR4). The dissertation contains both methodological developments and scientific applications, resulting in new photometric-redshift catalogues and new measurements of galaxy and quasar clustering.

The introductory chapters provide a comprehensive review of the theoretical foundations of large-scale structure formation and of the role played by redshift measurements in cosmological analyses. The discussion demonstrates a good understanding of both the cosmological context and the observational challenges associated with large photometric surveys.

A substantial part of the dissertation is devoted to machine-learning techniques for photometric-redshift determination. This section is particularly important because it forms the methodological foundation for all subsequent scientific results. The author reviews both traditional and modern approaches to photometric-redshift estimation, including template-fitting methods, supervised learning, artificial neural networks, and convolutional neural networks. The discussion is well organized and demonstrates a solid understanding of machine-learning principles such as training procedures, network optimization, bias-variance trade-offs, and feature extraction.

The central scientific achievement of the dissertation is the development of the Hybrid-z framework. The model combines four-band optical images with nine-band photometric magnitudes using a hybrid architecture consisting of convolutional neural networks and fully connected neural networks. Inception modules are employed to improve image feature extraction and to exploit morphological information contained in galaxy images. The work reported in this chapter has already resulted in a first-author publication in *Astronomy & Astrophysics*, which further confirms its scientific quality.

A particularly important aspect of this work is that Hybrid-z has been released as an open-source package and the resulting photometric-redshift catalogue has been made publicly available. This considerably enhances the scientific value and long-term impact of the contribution.

The first scientific application of Hybrid-z concerns the KiDS-DR4 bright galaxy sample. The author demonstrates that the inclusion of image-based information leads to a substantial improvement in photometric-redshift performance. The achieved reduction of the photo-z scatter by approximately 20% relative to the previous ANNz2 catalogue represents a significant advance. This result is not merely a technical improvement. Accurate redshift distributions are fundamental ingredients for all subsequent clustering analyses. Consequently, the development of the improved catalogue provides an important resource for future studies of the large-scale structure of the Universe.

The author subsequently extends the Hybrid-z methodology to quasars observed in the KiDS survey. This application is particularly interesting because photometric-redshift estimation for quasars is considerably more challenging than for galaxies due to their complex spectral properties and broad redshift range. The author successfully adapts the methodology to this problem and constructs redshift distributions suitable for clustering analyses. Using these improved photometric redshifts, the author measures the angular two-point correlation function and determines the evolution of quasar bias as a function of redshift. The results indicate a monotonic increase of bias from approximately 1.6 at redshift 0.6 to approximately 4 at redshift 2.2. The inferred halo masses are consistent with the current understanding of quasar environments and dark-matter halo occupation.

The final scientific component of the thesis investigates higher-order clustering statistics using the Count-in-Cells formalism. Moving beyond two-point statistics, the author studies non-Gaussian features of the galaxy distribution and their dependence on redshift, stellar mass, and color. The analysis confirms that red and massive galaxies exhibit stronger clustering than blue and less massive systems, consistent with contemporary models of galaxy formation and environmental evolution.

The dissertation presents original and scientifically significant research at the intersection of cosmology and machine learning. The work addresses an important problem in modern observational astronomy and provides a novel methodological solution through the development of the Hybrid-z framework.

The thesis demonstrates the candidate's ability to formulate scientific problems, develop advanced computational methodologies, analyze large observational datasets, and interpret the results within the framework of modern cosmology. The publication record associated with the dissertation further confirms its scientific quality and impact.

Therefore, I conclude that the presented dissertation meets the formal requirements for PhD theses and recommend admission of the Candidate to the subsequent stages of the procedure, including the public defense. In my opinion this PhD thesis is exceptional and deserves a special award.

Marek Demiański

Warsaw, June 9, 2026.