



**UNIVERSITÉ
DE GENÈVE**

FACULTÉ DES SCIENCES
Département d'astronomie

Prof. Stéphane PALTANI
Full professor

Tél : 022 379 21 49
Stephane.Paltani@unige.ch

Versoix, June 20, 2026

REVIEW OF THE DOCTORAL DISSERTATION:

**Photometric redshifts and clustering statistics of the
large-scale structure**

BY Anjitha John William

**SUBMITTED TO THE CENTER FOR THEORETICAL PHYSICS OF THE POLISH
ACADEMY OF SCIENCES**

This manuscript discusses several aspects of cosmology and galaxy evolution based on the Kilo-Degree Survey (KiDS) cosmological survey. The common thread is the use of photometric redshifts obtained with Hybrid-z, a deep-learning technique developed by the Candidate. Two applications of Hybrid-z are presented, on the bias of quasars and on count-in-cell statistics of the bright-galaxy population.

Part I very nicely introduces the cosmological context, starting from the Einstein field equation and its application to the expanding isotropic and homogeneous Universe. It then exposes the concept of redshift and the theory of linear growth of fluctuations. Finally, the clustering statistics that connect cosmology with observations and which will be used in the core of the dissertation are presented in detail. The introduction is well structured, and present all necessary ingredients for the understanding of the analyses conducted by the Candidate that will be part of the subsequent parts of this dissertation. Although there are a few inaccuracies here and there, it is in general well written, to the point and yet comprehensive enough.

Part II introduces the concept of photometric redshifts, which lies at the core of the work of the Candidate. After a quick introduction to the problem and to template-based algorithms, the manuscript focuses on machine-learning techniques, specifically deep learning. The presentation of deep-learning neural networks is an excellent technical introduction for scientists who are not familiar with the topic. However, for a part titled "Photometric redshift estimation", I feel that the manuscript glosses over a few important aspects of photometric-redshift estimation; for instance quality metrics, real-life examples, and state-of-the-art should have been presented here. Other aspects, such as the pros and cons of machine learning versus template fitting, the calibration of redshift distributions, or how machine-learning algorithms can be used in a probabilistic framework, are mentioned only very cursorily, without sufficient explanations. The latter

two aspects constitute actually a relative weakness that is present throughout the manuscript.

Part III is the Candidate's first paper, which presents Hybrid-z, a photometric-redshift algorithm developed by the Candidate. The algorithm combines standard artificial neural networks (ANNs) and convolutional neural networks (CNNs) to predict redshifts based on both the fluxes and the images in several photometric bands. Applied to bright galaxies in the KiDS survey, Hybrid-z improves the redshift predictions compared to a pure flux-based neural network approach with a reduction in the scatter of about 20%, which is fairly good. The development of the algorithm is sound, and a lot of different aspects have been considered. However, I feel some important context is missing on the astrophysical side. Firstly, at this level of signal-to-noise ratios, similar dispersions are expected from template fitting. Secondly, there have been many efforts to apply CNNs to the problem of redshift determination, even within the KiDS collaboration. The reason for additional work on the topic should have been better explained, especially since it is mentioned that other groups have obtained similar results. A specific question is whether the ANN part is needed or not. The combination of different predictors also implies a weighting scheme, but this was not discussed. In addition, the algorithm produces a single point estimate, without any uncertainties, which is usually not what the users want, as it prevents any necessary propagation of the uncertainties on the photometric redshift to the rest of the analysis. Finally, the discussion of the bias is largely insufficient; the complexity of the issue is demonstrated in the next Part.

Part IV is the second paper of the Candidate, and is currently under review. It addresses the determination of the bias of quasars in the KiDS survey. Hybrid-z is again used to determine the photometric redshifts. The clustering analysis is very well done in general, but the issue of the calibration of the photometric redshifts is an intrinsic limitation that is well identified by the Candidate. It is not clear to me if a proper calibration strategy suitable for quasars is workable, making spectroscopic surveys probably the most promising approach for studies of the clustering of quasars. The use of Hybrid-z should have also been better justified, as quasars are selected to be point-like. I would have liked to see a discussion of the adequacy of the CNN part of Hybrid-z for this application, and whether it leads to improved estimations of the photometric redshifts. The study and the discussion should have also been a bit more extensive. Important questions are the dependence of the bias on luminosity, and the properties of the halos hosting quasars.

Part V is work in progress. This is a very ambitious project that presents a count-in-cell analysis of the KiDS-DR4 galaxy sample based on photometric redshifts computed using Hybrid-z. The galaxy population is separated between blue and red galaxies. It is not clear to me to what extent the results that are presented are final, or whether the Candidate is still working on them. The upturn in the reduced cumulants is a clear example of unwanted features, which is adequately discussed by the Candidate. Effects of survey size, or other observational effects still need to be understood and modelled, for instance using a random sample with the same observational biases. As the Candidate mentions that other studies of the high-order statistics have been made in the KiDS collaboration, an in-depth comparison of the results is also in order. Having little experience in these high-order statistics myself, it is not clear to me how sensitive the measured parameters depend on an accurate photometric-calibration, but this should be discussed; my naive assumption is that the higher the order, the more sensitive to any biases. If correct, an additional calibration

stage is probably necessary. While it seems to me that a large amount of work is still needed, this work is quite innovative, and will make a significant contribution to the field.

In my view, the Candidate has demonstrated advanced skills in machine learning and in the methods of observational cosmology. It is clear that this dissertation represents a huge amount of work, and the expertise developed by the Candidate in the computation of clustering statistics will be very useful in any cosmological-survey collaborations. Therefore, I conclude that the presented dissertation meets the formal requirements for a Ph.D. thesis and recommend admission of the Candidate to the subsequent stages of the procedure, including the public defence.



Prof. Stéphane Paltani
Department of Astronomy
University of Geneva



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.)

June 20, 2026

Stéphane Pallani

.....

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)