Report on Doctoral Thesis by Suhani Gupta

B. F. Roukema
Institute of Astronomy
Faculty of Physics, Astronomy & Informatics
Nicolaus Copernicus University
Grudziadzka 5
87-100 Toruń

5 February 2024

I General assessment

The candidate presents her research on several aspects of modelling beyond-ΛCDM models. The models are those in which a flat, rigidly expanding background cosmological model is combined with structure formation, each of which is governed by a modified gravity model. As in ΛCDM, the two components of gravitational behaviour are assumed to be semi-decoupled, in the sense that gravity in the expanding background influences the role of gravity in structure formation, but expansion itself is unaffected by (semi-decoupled from) structure formation. The main body of the thesis is comprised of Chapter 2 on the halo mass function, Chapter 3 on the power spectrum, Chapter 4 on clustering and halo assembly bias, and Chapter 5 on other structure statistics.

Chapter 2 presents analytical 3-parameter ($f(R)$) and 4-parameter (nDGP) fitting functions of the deviation of the halo multiplicity function from that found numerically for ΛCDM, with parameters listed in Tables I and II of Gupta et al. (2022) and the functions in Eqs (2.13)–(2.15).

In Chapter 3, the candidate presents fitting functions for the $f(R)$ and nDGP models based on a combined 1-halo power spectrum and 2-halo power spectrum, termed "the halo model (HM)". The models start with the fitting models of Chapter 2 and add several new formulae and constants, summarised in Table I of Gupta et al. (2023). Although the computational reproducibility of the peer-reviewed research paper is severely limited, the data that *are* provided indicate confusion about which curve is which and what changes were made in the calculations (see §II.E). Nevertheless, giving the candidate the benefit of the doubt, the errors are unlikely to be major.

Chapter 4 presents unpublished work for the F5, F6, N1 and N2 models for halo assembly bias related characteristics, including halo bias in Fig. 4.1, halo concentration in Fig. 4.2, halo spin in Fig. 4.3. Overall, the candidate shows that the four models are only weakly distinguished from ΛCDM by these characteristics.

In Chapter 5, the candidate compares the F5, F6, N1 and N2 models to ΛCDM in terms of the probability density function of morphological characteristics of the flat-space gravitational field as determined by the eigenvalues of the Hessian $\partial^2\phi(\vec{x},t)/\partial x_\alpha \partial x_\beta$; density field moments; morphological characteristic halo mass functions, halo multiplicity functions, and halo spins. The results found were that most of these characteristics are sensitive to the choice of model.

The candidate shows reasonable familiarity with the standard model of cosmology, as is generally accepted for doctorates in cosmology. The focus on alternative gravity models without specifying that these share the semi-decoupled gravity model of ΛCDM is a significant flaw in the motivation. While the semi-decoupling hypothesis is popular, it is nevertheless a hypothesis that is absent from the Einstein equation. A more conservative alternative to ΛCDM than $f(R)$ and $nDGP$ models is to drop the semi-decoupling hypothesis and see if the standard Einstein equation is sufficient to model the observations (Buchert et al., 2015; Lapi et al., 2023; Carfora & Familiari, 2024).

The thesis is generally well-presented, with a coherent and thorough approach to a useful question in current extragalactic astronomy, independently of the motivation, demonstrating the candidate's scientific independence and constituting original scientific work, including a primary role in two peer-reviewed papers in a prestigious journal.

The modelling and the results themselves are not easily reproducible in the state-of-the-art sense of Akhlaghi et al. (2021). The file provided as a small step towards reproducibility shows only rough consistency with the corresponding published figure and suggests errors in labelling curves. However, despite the increasing intensity of requests for reproducibility from funding agencies such as NCN, computational reproducibility is still an emerging standard of science that is currently only *recommended*, not *required* in a strict sense. Several months of work would likely be needed for a cosmologist to reproduce results similar to those of the candidate prior to being able to carry out followup research (see §II.G), but the descriptions provide enough clues, within the bounds of the currently accepted standards of the international astronomy community.

Overall this is a good PhD thesis presenting a substantial body of original research in extragalactic numerical modelling.

There are several minor concerns; some concerning more fundamental aspects of scientific reasoning, and some concerning the clarity of the presentation. These are listed as follows.

II Concerns

II.A Abstract

1. The jargon "systematics" would have better been written "systematic errors", to avoid adding unnecessary obstacles to scientists from other fields wishing to understand the thesis.

II.B Table 1, generic minor concerns

1. Writing "cric" rather than "crit" for the subscript of the critical density is a bit confusing, since it suggests that is different to the usual critical density. Moreover, since the subscript does not represent $c \times r \times i \times c$, it should be in roman font (e.g. `\mathrm{crit}`).

2. Units, the "d" for differentials, and function names should also be in roman font, e.g. Mpc, $\mathrm{d}s^2$, sin. This standard of style aids in clarity; again, to avoid confusing, e.g. $s \times i \times n$ with sin.

3. The distance unit "Mpc" should not have a space between "M" and "pc".

4. There are small aspects of grammar and style or minor misuses of words throughout the thesis (such as capitalisation; "in-turn" instead of "in turn") that are generally contrary to what is expected in a good journal; these will normally be caught at the proof-reading stage.

II.C Chapter 1. Introduction

1. The candidate states that ΛCDM is "based on GR", e.g. "The standard cosmological paradigm, which is based on GR, is referred to as ΛCDM", without clarifying that the ΛCDM model is relativistically a semi-decoupled model: expansion affects structure formation, but structure formation is forbidden from affecting expansion in ΛCDM. It is misleading to state that ΛCDM is "based on GR" without specifying that it is a model that is only partly relativistic.

   This is especially relevant in the introductory section of a thesis that aims to consider beyond-GR models rather than take the more conservative approach of considering GR models that are beyond-ΛCDM.

2. In Eq. (1.2), the candidate makes no comment on the singularity at the equator in the spherical case in this expression: when $1 = Kr^2$ in the second term, there is a division by zero, which is undefined in arithmetic on the real numbers. While this is common in cosmology textbooks, a doctoral candidate should check the validity of background material, especially when division by zero occurs.

3. Eq. (1.3) and its description are wrong, except for the special case of a flat FLRW universe. Curved spaces are not vector spaces. The candidate gives no explanation of how FLRW models with spatial sections $\mathbb{S}^3$, $\mathbb{S}^3/\mathbb{Z}_2$, $\mathbb{S}^3/I^*$, or $\mathbb{H}^3$ are to be represented as a vector space, nor is there an explanation provided about the consequences of choosing an appropriate projection from a curved space to a flat space.

   Moreover, galaxy formation is forbidden at the turnaround epoch in a flat model (Roukema & Ostrowski, 2019), which the candidate fails to explain. Since the candidate primarily considers large-scale structure as traced by galaxy haloes, this contradiction should have at least been commented on.

4. page 5, "GR gives prescription ... $a \propto e^t$ showing exponential expansion of our Universe". Trivially, the expression should be written $a \propto \mathrm{e}^t$, since e is a mathematical constant, not a variable. More importantly, the difference between an approximation and an exact expression is fundamental in the physical sciences; this is neglected in the wording chosen by the candidate.

5. "A plethora of observational evidences indicate that it is indeed the case, and our Universe is flat (i.e. $\Omega_K = 0$)" misrepresents the evidence presented in the papers cited. The papers cited show that the best fitting FLRW model to the data is very close to flat on the scale of the observational sphere, but no evidence that the Universe itself is exactly flat. ΛCDM interpreted in the inflationary context is a model in which the curvature of the Universe as a whole is completely unknown, masked by inflation and the tiny size of the observable sphere. Evidence for the Universe being exactly flat in the geometrical sense would be possible if the Universe were shown to be spatially a $\mathbb{T}^3$ model, for example (Roukema & Edge, 1997; Akrami et al., 2022, e.g.). The difference between a fitting function and reality is a fundamental difference.

6. The candidate writes, "a directly related effect to the expansion is that the physical wavelength of light emitted from a distant object is stretched out proportionally to $a(t)$". This is acceptable

as a memorisation trick in cosmology classes[1] to help remember the relation, but is physically misleading (Narlikar, 1994; Synge, 1964).

7. The first paragraph of §1.2 is meaningless without first clarifying the restriction to an exactly flat spatial section. Representing either a globally non-flat spatial section or an approximately flat spatial section (in which structure formation is allowed to induce non-zero spatial curvature) by a vector space requires a clear definition on how this is to be done.

8. Eqs (1.8)–(1.11) are only valid in the flat case. At least a brief reminder should have been provided to the reader about this caveat.

9. "First, none of the present theories can incorporate both GR and the quantum theory." On the contrary: there exist several families of approaches to quantum gravity that "incorporate" both GR and quantum mechanics: string theory, loop quantum gravity and causal dynamical triangulation are among the most popular. None are currently accepted as experimentally supported, but the theories certainly exist as active fields of research.

10. page 15, "so-called Cosmological Constant": I would recommend that the candidate remove the use of "so-called" in future scientific texts, except where sarcasm is specifically intended (which is rarely acceptable in a scientific text).

11. page 26: It is misleading to describe RAMSES as only being "publicly available". What is significant about the code is that it is FOSS (free-and-open-source-software), licensed under CECILL[2], under a licence that obliges modified versions to be compatibly licensed. The candidate's version of the code necessarily gives her the same freedoms to use, modify, distribute or distribute modified copies (at zero cost or for a fee).

12. page 26: "simulation are run" has a typo; "MPGRapfic [232]," is misspelt.

II.D Chapter 2. Halo Mass Function in Modified Gravity cosmologies

1. page 34: "The authors in [24] showed that the HMF varies with the spectral index of the primordial power spectrum," is true in describing Bagla+2009, but Eq. (2.6) shows that the statement was already shown in 1974, 35 years earlier.

2. page 43: "theory in-still a high level" – typo "instill"

II.E Chapter 3. Analytical modelling of the Power Spectrum in Modified Gravity cosmologies

1. This chapter begins with a discussion of the cosmological power spectrum and its relevance without reminding the reader that this is only defined in exactly flat space, nor that the spectrum is discrete unless space is infinite. However, the spatial sections of general relativistic spaces are, in general, curved, and there is no observational evidence establishing the size of the spatial section of a general relativistic universe to be infinite. This is a problem of missing context.

[1] See https://en.wikipedia.org/wiki/Lies_for_children.
[2] https://www.cecill.info/index.en.html

2. page 61, §3.1, "where the lensing signal is sourced by the distribution of all matter in the Universe" – lensing effects only occur along null geodesics – straight lines – between background sources and the observer. Moreover, if ΛCDM is interpreted to have a spatially infinite spatial section, then "all matter in the Universe" implies that an infinite volume of matter has effects on observations in a single observable sphere. This wording is confusing.

3. Fig 3.1 – it is unclear what the "shaded region" refers to.

4. Since the candidate and her co-authors have made some of their data available, comparing Figs 2a–2d of Gupta et al. (2023) to the data files published is straightforward. Exact instructions for doing this are listed in Appendix IV of this report, and the figures are shown below in Fig. 1 of this report.

   While the file `Description.pdf` says that the numbers refer to the halo models, the corresponding curves in the published paper (thesis pdf page 115; journal page 083425-8) that most closely match the data file values appear to be a mix of those marked "Simulation" and "Halo Model" in the published paper, with noticeable variations in values.

   (a) *(top-left):* The $\Upsilon(k = 7\,h/\mathrm{Mpc}) \approx 1.030$ file value for F6 and the curve in Fig. 1 top-left (dotted) nearly match the "Simulations" curve of the published Fig 2a, which has $\Upsilon(k = 7\,h/\mathrm{Mpc}) \approx 1.05$ even though the shape of the curve differs between the published figure and the plain text data; the F5 file curve may match part of the published curve, except for $k \gtrsim 2\,h/\mathrm{Mpc}$.

   (b) *(top-right):* The N1 "Simulations" curve for $z = 0$ in the published figure has $\Upsilon(k = 7\,h/\mathrm{Mpc}) \approx 1.05$ and the N1 "Halo Model" curve has $\Upsilon(k = 7\,h/\mathrm{Mpc}) \approx 0.98$ versus $\Upsilon(k = 7h/\mathrm{Mpc}) \approx 1.017$ in the plain text file (solid curve in this report).

   (c) *(bottom-left):* For $z = 0.5$, the numbers for F5 and F6 could almost match the "Halo Model" curves plotted in the published paper, except for $k \gtrsim 3\,h/\mathrm{Mpc}$ for F5 where the shapes are clearly different;

   (d) *(bottom-right):* For $z = 0.5$, the data file numbers for N1 and N2 appear to nearly match the "Halo Model" curves, except that $\Upsilon(k = 7h/\mathrm{Mpc}) \approx 1.02$ for N1 in the published figure, compared to $\Upsilon(k = 7.02h/\mathrm{Mpc}) = 1.065$ for N1 in the data file; and the baryon acoustic oscillations are absent in the published figure.

   The published paper states "The data used here is publicly available on our website" regarding the `.tgz` file, while "Description.pdf" states that the contents are "based on" the published paper.

   These inconsistencies suggest only annoying, not major, errors. However, they do constitute obstacles to others aiming to verify the work.

5. Appendix B of Gupta et al. (2023) states that the candidate and co-authors used a software package called `HMcode`, which is proprietary[3]. The `HMcode` package is not open-source[4], since the lack of an explicit copyright statement puts it under the default copyright, i.e. the Berne Convention. While the source code is available for inspection, permission is *not* given: to sell

[3] swh:1:rev:79109e37d77be7ea07aa7f3ea6d80406c5af52fc

[4] `https://en.wikipedia.org/wiki/The_Open_Source_Definition`

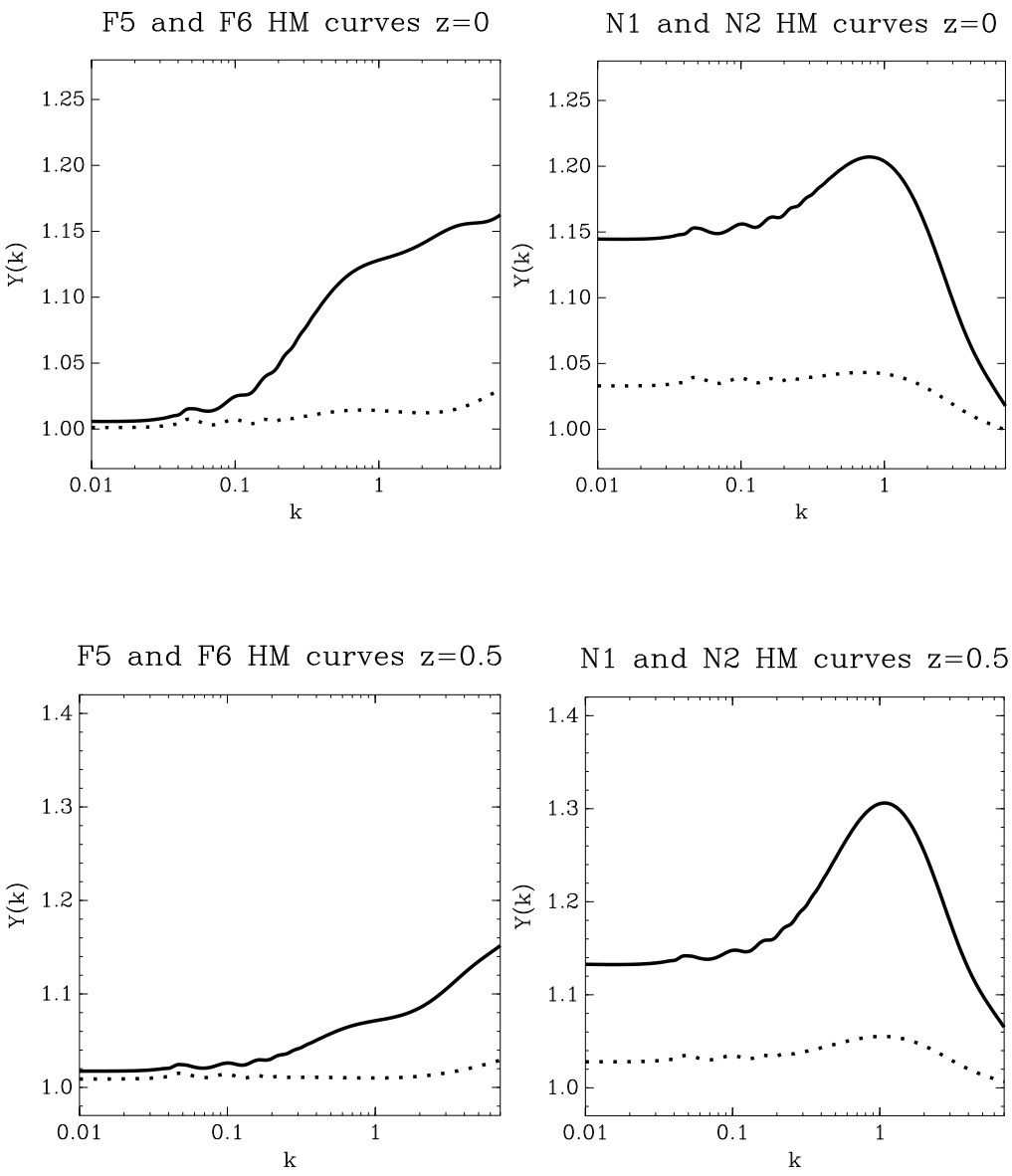

Figure 1: $\Upsilon(k)$ for the *Halo Models* for: *(top-left):* the $z = 0$, F5 (solid) and F6 (dotted) models; *(top-right):* the $z = 0$, N1 (solid) and N2 (dotted) models; *(bottom-left):* the $z = 0.5$, F5 (solid) and F6 (dotted) models; *(bottom-right):* the $z = 0.5$, N1 (solid) and N2 (dotted) models, based on "Description.pdf" and the plain text files of `https://web.archive.org/web/20240203191714/https://data.cft.edu.pl/UPSILON_PK/UpsilonPk.tar.gz` . See Appendix IV for the script that produces these plots.

or give away the software (violating criterion 1), to distribute the original source or compiled code (violating criterion 2), to modify and distribute the modified software (violating criterion 3), for someone who buys or is given the software to redistribute it further without a new licence (violating criterion 7). Thus, `HMcode` is *not* open-source software; those who use it or redistribute it are potentially at risk of legal challenges for copyright violation, especially if the redistribution is international.

II.F Chapter 5. Modified Gravity in the Cosmic Web

1. page 115, §5.1 "delve more information" probably means "provide more information".

2. page 115, §5.1, sentences should not start with "And".

II.G Chapter 6. Summary and future prospects

1. page 140, para 2, "alongr"

2. The analysis does not satisfy state-of-the-art criteria for reproducible research (Akhlaghi et al., 2021) and open quantitative science[5]. For example, the reader is not informed about checksum-identified URLs (or more generically, URIs) of the input data files, and the precise versions of analysis software source code, software libraries, the detailed methods of compilation, and scripts for doing the analysis, are not known to the reader. In practical terms, another cosmologist than the candidate would require weeks or longer to reproduce roughly compatible results, or maybe months if the attempted verification starts in 2034 instead of immediately. A reproducible research project should require typically a few hours or a day for full, detailed numerical verification by an independent cosmologist. Full documentation of data and software lineage and archiving of these are strongly encouraged by grant-giving organisations in Poland and much of Europe.

3. The candidate should consider doing the followup projects in line with state-of-the-art reproducibility criteria, as per the work of other recent doctoral students (Peper & Roukema, 2021; Borkowska & Roukema, 2022; Peper et al., 2023). Moreover, applications for grants will require that the candidate describes plans for reproducibility.

II.H Other minor flaws

1. The candidate should remove the sarcastic term "so-called" throughout the thesis. (Alternatively, use of the adjective should, at least, be accompanied by a justification of why the candidate finds a piece of terminology dubious.)

2. There are numerous minor flaws in the English, but generally these do not prevent comprehension.

[5] `https://sorbonnedatadeclaration.eu`

III Recommendation

Gupta's PhD thesis satisfies the current international standards of scientific research in modern cosmology expected for a doctoral thesis as well as the formal requirements for obtaining the degree. I recommend that the candidate be admitted to the subsequent stages of the thesis procedure, and, subject to satisfactory performance in the defence, be awarded a doctoral degree.

prof. dr hab. Boudewijn F. Roukema
5 February 2024, Toruń

References

Akhlaghi M., Infante-Sainz R., Roukema B. F., Valls-Gabaud D., Baena-Gallé R., 2021, Comp. in Sci. Eng., 23, 82 (arXiv:2006.03018)

Akrami Y., et al., 2022, arXiv e-prints, (arXiv:2210.11426)

Borkowska J., Roukema B. F., 2022, Classical and Quantum Gravity, 39, 215007 (arXiv:2112.14174)

Buchert T., et al., 2015, Classical and Quantum Gravity, 32, 215021 (arXiv:1505.07800)

Carfora M., Familiari F., 2024, arXiv e-prints, (arXiv:2401.04293)

Gupta S., Hellwing W. A., Bilicki M., García-Farieta J. E., 2022, Physical Review D, 105, 043538 (arXiv:2112.03699)

Gupta S., Hellwing W. A., Bilicki M., 2023, Physical Review D, 107, 083525 (arXiv:2301.12016)

Lapi A., Boco L., Cueli M. M., Haridasu B. S., Ronconi T., Baccigalupi C., Danese L., 2023, Astrophys.J., 959, 83 (arXiv:2310.06028)

Narlikar J. V., 1994, American Journal of Physics, 62, 903

Peper M., Roukema B. F., 2021, MNRAS, 505, 1223 (arXiv:2010.03742)

Peper M., Roukema B. F., Bolejko K., 2023, MNRAS, 525, 91 (arXiv:2304.00591)

Roukema B. F., Edge A. C., 1997, MNRAS, 292, 105 (arXiv:astro-ph/9706166)

Roukema B. F., Ostrowski J. J., 2019, JCAP, 12, 049 (arXiv:1902.09064)

Synge J. L., 1964. Amsterdam: North-Holland

IV Appendix

The following commands should produce plots with `plotutils`[6] 2.6-13. These are given here so that the reader of this review can verify that the figures in this report match the source data file, of which the `.tar.gz` file has an `sha256sum` value of `10692e2476dc0d5bfd41b52943816231c53dbc5345d7cc5d356a6033ac48e66d` .

```
  wget https://web.archive.org/web/20240203191714\
/https://data.cft.edu.pl/UPSILON_PK/UpsilonPk.tar.gz
  tar -xv -f UpsilonPk.tar.gz
  cd UPSILON_PK
  for n in 5 6; do grep -v ^# UPSILON_F${n}/upsilon_pk_f${n}_z0.00.dat;
        printf "\n"; done | \
          graph -Tps -lx -FHersheySerif -X k -Y "\\*U(k)" \
            -L "F5 and F6 HM curves z=0" -W 0.005 \
            -x 0.01 7 -y 0.97 1.28 > Gupta2023Fig2a.eps
  for n in 1 5; do grep -v ^# UPSILON_N${n}/upsilon_pk_n${n}_z0.00.dat;
        printf "\n"; done | \
          graph -Tps -lx -FHersheySerif -X k -Y "\\*U(k)" \
            -L "N1 and N2 HM curves z=0" -W 0.005 \
            -x 0.01 7 -y 0.97 1.28 > Gupta2023Fig2b.eps
  for n in 5 6; do grep -v ^# UPSILON_F${n}/upsilon_pk_f${n}_z0.50.dat;
        printf "\n"; done | \
          graph -Tps -lx -FHersheySerif -X k -Y "\\*U(k)" \
            -L "F5 and F6 HM curves z=0.5" -W 0.005 \
            -x 0.01 7 -y 0.97 1.42 > Gupta2023Fig2c.eps
  for n in 1 5; do grep -v ^# UPSILON_N${n}/upsilon_pk_n${n}_z0.50.dat;
        printf "\n"; done | \
          graph -Tps -lx -FHersheySerif -X k -Y "\\*U(k)" \
            -L "N1 and N2 HM curves z=0.5" -W 0.005 \
            -x 0.01 7 -y 0.97 1.42 > Gupta2023Fig2d.eps
```

[6] https://www.gnu.org/software/plotutils or https://tracker.debian.org/pkg/plotutils; e.g. `apt install plotutils`

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)

„Linear and non-linear statistics of the cosmic density field in modified gravity cosmologies”.

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

Suhani Gupta

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

Pozytywna ocena *(Positive conclusion)*:

[X] 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.)

PODPIS ZAUFANY
BOUDEWIJN
ROUKEMA
05.02.2024 15:11:25 [GMT+1]
Dokument podpisany elektronicznie
podpisem zaufanym

.....
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)